

Université de Montréal

Modélisation d'un réseau de neurones humains dans le but de
comprendre la dégradation neurale lors du vieillissement

par

Rémy Allard

École d'optométrie

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de Maîtrise (M.Sc.)

en sciences de la vision

option sciences fondamentales

Septembre 2003

© Rémy Allard, 2003



ww

5

U58

2004

v.003

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
Modélisation d'un réseau de neurones humains dans le but de comprendre la
dégradation neurale lors du vieillissement

présenté par :
Rémy Allard

a été évalué par un jury composé des personnes suivantes :

Christian Casanova, Ph.D.
président rapporteur

Jocelyn Faubert, Ph.D.
directeur de recherche

Langis Michaud, O.D., M.Sc.
membre du jury

RÉSUMÉ

Pendant plusieurs années, la croyance était que le cerveau, incluant le cortex visuel, perdait des neurones en vieillissant et que ces pertes de neurones causaient des déficits cognitifs et perceptifs. Des études plus récentes ne rapportent aucune perte de neurones avec le vieillissement pour la plupart des aires corticales incluant le cortex visuel. Ces études ont ouvert la porte à d'autres portant sur les altérations neurobiologiques (ANBs) pouvant corrélérer avec les pertes cognitives et perceptives dues au vieillissement chez l'humain.

Les réseaux de neurones artificiels (RNA) sont de bons modèles pour simuler le fonctionnement du cortex et son vieillissement. La première étude se consacra à la création d'un RNA capable d'apprendre des tâches visuelles simples et complexes. Tout d'abord, il fallut développer une méthode adaptative qui ajuste le taux d'apprentissage et facilite ainsi l'apprentissage de tâches complexes pour des RNA étendus. La tâche visuelle simple consistait à détecter un sinus modulé par la luminance (premier ordre). La tâche complexe consistait à détecter un sinus modulé par le bruit (deuxième ordre).

Avec le vieillissement, les tâches plus complexes semblent plus affectées que les tâches simples. Dans la deuxième étude, des ANBs ont été simulées dans le but

d'identifier celles qui corrèlent avec les déficits visuels perceptifs. Plusieurs ANBs telles la perte de synapses, la dégradation de la gaine de myéline et la dégénérescence de l'inhibition intra-corticale, ont démontré une bonne corrélation avec les données psychophysiques sur le vieillissement. Ces données correspondent à la performance d'un groupe de jeunes adultes et d'un groupe de personnes âgées pour des tâches visuelles simples et complexes. Habak et Faubert (2000) ont démontré que les seuils de détections des personnes âgées augmentent de 29 % et de 77 % pour des stimuli de premier et de deuxième ordres respectivement, en comparaison aux seuils des jeunes. Les simulations ont démontré que plusieurs hypothèses d'ANBs corrèlaient avec ces résultats. Cependant, ces différentes hypothèses ne prédisent pas les mêmes performances pour des groupes d'âge non testés par Habak et Faubert (2000). Pour les personnes de moins de 70.75 ans, certaines hypothèses prédisent que le ratio d'augmentation des seuils de premier et de deuxième ordres serait constant alors qu'une autre prédit que le deuxième ordre serait davantage affecté. Pour les sujets très âgés, certaines hypothèses prévoient également un ratio constant alors que d'autres prédisent que le premier ordre serait davantage affecté.

Connaître la progression du ratio, due au vieillissement, des déficits de premier et deuxième ordre permettrait aux simulations effectuées dans ce présent mémoire d'appuyer ou non certaines hypothèses d'ANBs comme étant une cause importante des déficits perceptifs.

Mots clés

Réseau de neurones artificiels; Apprentissage; Backpropagation; Méthode adaptative; Taux d'apprentissage; Simulation; Vieillissement; Premier ordre; Deuxième ordre; Perception; Cortex visuel

SUMMARY

During several years, it was believed that the aging brain, including the visual cortex, lost neurons and that these neurons losses caused cognitive and perceptive deficits. Recent studies, however, do not report any neuron losses due to aging for the majority of the cortical surfaces, including the visual cortex. These studies have opened the door to other studies on neurobiological alterations (NBAs) that correlate with the cognitive and perceptive losses due to human aging.

Artificial neural networks (ANNs) are good models to simulate the cortex and the aging cortex. The first study was devoted to the creation of an ANN able to learn simple and complex visual tasks. In order to achieve this, it was first necessary to develop an adaptive method that adjusts the learning rate and thereby facilitates the complex task learning of broad ANNs. The simple visual task consisted in detecting a sine modulated by brightness (first order). The complex task consisted in detecting a sine modulated by the noise (second order).

With aging, complex tasks seem to be more affected than simple tasks. In the second study, NBAs were simulated with the goal of identifying those that correlated with the perceptive visual deficits. Several NBAs such as synapse loss, myelin sheath degradation and degeneration of intra-cortical inhibition, showed a

good correlation with the psychophysical data on aging. This data corresponds to the performance of a group of young adults and a group of elderly for a simple and a complex visual tasks. Habak and Faubert (2000) showed that the thresholds of detections of elderly increase by 29 % and 77 % for stimuli of first and second order respectively, this in comparison with the thresholds of the young adults. Simulations showed that several NBAs hypotheses correlated with these results. However, these various hypotheses do not predict the same performances for other age groups that were not tested by Habak and Faubert (2000). For adults younger than 70.75 years-old, certain hypotheses predicted that the threshold increase ratio of first and second order would be constant and others predicted that the second order would be more affected. For subjects over 70.75, certain hypotheses also predicted a constant ratio and others predicted that the first order would be more affected.

To know the progression ratio, due to ageing, of first and second order deficits would allow the simulations carried out in this present report to support or not certain NBA hypotheses as being a significant cause of the perceptive deficits.

Keywords

Artificial neural network; Learning; Backpropagation; Adaptive method; Learning rate; Simulation; Aging; First order; Second order; Perception; Visual cortex

TABLE DES MATIÈRES

Résumé	i
Summary.....	iv
Table des matières	vi
Liste des tableaux et des figures.....	ix
Liste des sigles et des abréviations.....	x

Chapitre I

Introduction	1
Recension de la littérature	3
1. Le vieillissement perceptif.....	3
2. Vieillissement et comportement neural	7
3. Base neurale du vieillissement.....	9
4. Les réseaux de neurones artificiels	14
5. Modélisation d'un réseau de neurones biologiques	20
Contexte de l'étude	24

Chapitre II

Article 1	25
Abstract.....	26
Introduction.....	27

The texture task.....	28
First experiment: an exhaustive search for the optimal learning rate	31
Second experiment: the naïve adaptive method	40
Third experiment: the no-decrease adaptive method.....	43
Fourth experiment: the generalized no-decrease adaptive method.....	52
General discussion	56
References.....	58
Annex: pseudo-code of the GNDAM	60

Chapitre III

Article 2	64
Abstract.....	65
Introduction.....	66
Methodology.....	71
Results	78
Discussion.....	83
References.....	88

Chapitre IV

Discussion générale des résultats	92
---	----

Chapitre V

Conclusion.....	97
Bibliographie.....	99

LISTE DES TABLEAUX ET DES FIGURES

Chapitre II

Article 1 : The generalized no-decrease adaptive method for large-slow-learning neural networks

Figure 1 : First and second order stimuli	29
Figure 2 : Small second order stimuli	32
Figure 3 : Backpropagation: error rate	36
Figure 4 : Backpropagation: error rate evolution.....	37
Figure 5 : Backpropagation: error rate evolution of the optimal fix learning rate.	38
Figure 6 : NDAM: learning rate evolution	50
Figure 7 : GNDAM: learning rate evolution.....	53
Figure 8 : GNDAM: error rate evolution	54

Chapitre III

Article 2 : Simulating age-related neurobiological dysfunctions with artificial neural networks: Effect on simple and “Gestalt-like” visual perception

Figure 1 : First and second order stimuli	69
Figure 2 : Artificial neural network architecture	73
Figure 3 : First and second order error rate.....	78
Table I : Neurobiological alteration simulation results.....	80
Figure 4 : Aging curve hypotheses	84

LISTE DES SIGLES ET DES ABRÉVIATIONS

AM : *adaptive method*

ANB : *alteration neurobiologique*

BP : *backpropagation*

GNDAM : *generalized no-decrease adaptive method*

LR : *learning rate*

NAM : *naïve adaptive method*

NBA : *neurobiological alteration*

NDAM : *no-decrease adaptive method*

OAPER : *optimal a priori error rate*

pdf : *probability distribution function*

RNA : *réseau de neurones artificiels*

Chapitre I

INTRODUCTION

Les réseaux de neurones artificiels (RNAs) peuvent aider à la compréhension du vieillissement neural chez l'humain. Le présent mémoire vise à mieux déterminer les causes neurales des pertes perceptives occasionnées par le vieillissement. Pour ce faire, des simulations artificielles d'altérations neurobiologiques (ANBs) pouvant expliquer les pertes perceptives dues au vieillissement ont été réalisées.

Avant de pouvoir simuler des ANBs, un RNA devait préalablement apprendre à effectuer certaines tâches : une tâche de perception de textures simples (premier ordre) et une de textures complexes (deuxième ordre). Des difficultés se sont présentées pour apprendre à un RNA une tâche de deuxième ordre. Le premier article présente donc un nouvel algorithme d'apprentissage utile à ce type de tâche.

La deuxième étude consiste à simuler artificiellement le vieillissement pour des tâches perceptives visuelles de premier et de deuxième ordres. Elle démontre que plusieurs ANBs corrèlent avec les déficits perceptifs occasionnés par le vieillissement notés dans la littérature, mais que ceux-ci ne prédisent pas les mêmes déficits pour différents groupes d'âge.

Afin de familiariser le lecteur, une recension de la littérature précède les deux articles. La première partie décrit les déficits perceptifs subis lors du vieillissement. La deuxième section définit deux altérations du comportement neural qui surviennent lors du vieillissement. En troisième lieu, les différentes ANBs causées par le vieillissement sont expliquées. La quatrième section présente une introduction aux RNAs ainsi qu'une revue des principaux algorithmes d'apprentissage connus dans la littérature. Enfin, la dernière partie de la recension de la littérature introduit des simulations d'ANBs qui sont effectuées dans la deuxième étude.

L'ouvrage se conclut par une discussion et propose de futures avenues de recherche.

RECENSION DE LA LITTÉRATURE

1. LE VIEILLISSEMENT PERCEPTIF

La majorité des fonctions perceptives atteignent leur maximum d'efficacité durant l'adolescence et maintiennent généralement ce pic pour plus d'une vingtaine d'années. Après la quarantaine, certains déficits perceptifs commencent à être notables. Plusieurs études récentes ont tenté de comprendre les pertes perceptives causées par le vieillissement (Faubert, 2002). La différence entre les performances d'un groupe de sujets adultes et d'un groupe de sujets âgés est généralement utilisée pour évaluer les changements occasionnés par le vieillissement. Pour s'assurer que ces pertes sont causées par un déficit perceptif et non cognitif de plus haut niveau, les chercheurs utilisent généralement des tâches très simples comme identifier l'orientation, verticale ou horizontale, ou encore identifier la direction de mouvement, gauche ou droite, de barres de contraste varié. Des tâches plus complexes auraient pour effet de soulever des questions sur l'origine, cognitive ou perceptive, des déficits. Séparer les déficits perceptifs de ceux sensoriels est plus complexe puisque tout traitement perceptif se base sur l'information sensorielle. Lorsqu'un déficit perceptif est noté, la question de savoir s'il est causé par un déficit sensoriel est donc importante.

Pour s'assurer qu'un déficit noté lors du vieillissement n'est pas causé par une perte sensorielle, le chercheur peut, par exemple, simuler une perte sensorielle chez les jeunes similaires à celle notée chez les personnes âgées (ex : filtre de densité neutre qui simule l'opacité optique causée, entre autres, par la perte de transparence du cristallin) ou, pour chaque sujet, ajuster la visibilité des stimuli en fonction de leur propre seuil de détection. Par exemple, si une étude vise à comprendre l'effet du vieillissement sur la perception du mouvement, il est important de s'assurer que le déficit enregistré n'est pas causé par une diminution de la sensibilité au contraste qui se produit également lors du vieillissement. Il est donc possible d'ajuster le contraste des stimuli de sorte que leur visibilité soit équivalente pour les jeunes et pour les personnes âgées.

1.1 Hypothèse prédicatrice des déficits perceptifs

Après plusieurs études sur le vieillissement perceptif, Faubert (2002) conclut que plus une tâche est complexe ou plus une tâche nécessite un réseau d'intégration neural étendu, plus un déficit causé par le vieillissement sera prononcé. En d'autres termes, toutes tâches de haut niveau comme la perception de la symétrie, l'intégration entre attributs ou l'intégration spatio-temporelle sont affectées par le vieillissement puisqu'elles requièrent un réseau de neurones étendu ou complexe (Faubert, 2002).

1.2 Stimuli de premier et de deuxième ordres

La sensibilité est définie comme étant l'inverse du seuil de détection. Le seuil de détection est le contraste minimal nécessaire pour percevoir un stimulus donné. Les stimuli de premier ordre sont définis, pour le présent mémoire, par un contraste de luminance, alors que ceux du deuxième ordre sont définis par un contraste de textures. Un stimulus de premier ordre est une alternance progressive de bandes pâles et foncées (réseau sinusoïdal) dans du bruit. Les stimuli de deuxième ordre sont également formés de réseaux sinusoïdaux sauf que ces réseaux sont modulés dans le bruit. La luminance de ce bruit est, en moyenne, équivalente pour tout le stimulus sans quoi il y aurait un indice de premier ordre. Dans les zones de haut contraste, le bruit varie largement c'est-à-dire que la luminosité varie aléatoirement entre le blanc et le noir mais conserve une luminosité moyenne constante. Par contre, dans les zones de bas contraste, le bruit varie seulement entre du gris pâle et du gris foncé. Un stimulus de deuxième ordre demande donc un traitement supplémentaire comparativement à un stimulus de premier ordre (Cavanagh & Mather, 1989; Wilson, 1998).

En effet, pour percevoir le réseau sinusoïdal d'un stimulus de premier ordre, il suffit de comparer les différentes luminosités du sinus. Par contre, pour percevoir le réseau sinusoïdal d'un stimulus de deuxième ordre, une comparaison local (au niveau du bruit) doit préalablement être faite pour déterminer le niveau de variance local du bruit, puis une comparaison entre les différentes variations de bruit permet

alors de percevoir le sinus. Les stimuli de premier ordre peuvent être définis mathématiquement par la fonction suivante :

$$L(x, y) = 0.5 \cdot \text{rand}(x, y) + 0.25 + 0.25 \cdot \text{Amp} \cdot \sin\left(\frac{2 \cdot \pi \cdot z}{p}\right)$$

$L(x, y)$ représente la luminance normalisée (entre 0 et 1) du pixel à la position (x, y) . $\text{rand}(x, y)$ retourne une valeur aléatoire entre 0 et 1 pour chaque pixel. Amp est l'amplitude de contraste choisie aléatoirement entre 0 et 1 pour chaque stimulus. p est la période du sinus et z dépend de l'orientation ($=y$ pour horizontale et $=x$ pour verticale).

Les stimuli de deuxième ordre peuvent être définis mathématiquement par la fonction suivante :

$$L(x, y) = 0.5 + 0.5 \cdot \text{Amp} \cdot \text{rand}(x, y) \cdot \sin\left(\frac{2 \cdot \pi \cdot z}{p}\right)$$

Noter que les stimuli de premier et de deuxième ordres ont toujours la même luminance moyenne et la même variation moyenne de bruit. La différence est que pour le premier ordre la variance du bruit est toujours constante et la luminance est modulée par le sinus alors que pour le deuxième ordre la luminance est toujours constante et la variance du bruit est modulée par le sinus.

1.3 Vieillessement et perception de stimuli de premier et deuxième ordres

Habak et Faubert (2000) ont comparé le contraste de luminance (premier ordre) et le contraste de bruit (deuxième ordre) nécessaires à la perception d'un stimulus chez de jeunes adultes (âge moyen de 22.75 ans) et des personnes âgées (âge moyen de 70.75 ans). Ils ont trouvé que les seuils de détection des personnes âgées comparativement aux jeunes augmentaient de 29 % et de 77 % pour le premier ordre et le deuxième ordre respectivement. Ces résultats vont dans le même sens que l'hypothèse stipulant que les tâches plus complexes sont plus affectées en vieillissant (Faubert, 2002).

2 VIEILLISSEMENT ET COMPORTEMENT NEURAL

2.1 Sollicitation de réseaux de neurones plus étendus avec le vieillissement

Des études, à l'aide d'imagerie par résonance magnétique fonctionnelle, ont démontré que chez les personnes âgées, un stimulus analysé par le cortex visuel active un réseau de neurones plus étendus que chez les jeunes (Bennet, Sekuler, McIntosh, & Della-Maggiore, 2001; Cabeza, 1997; Madden, Turkington, Provenzale, Denny, Hawk, Gottlob, & Coleman, 1999). Faubert (2002) propose comme explication qu'il y aurait une moins grande efficacité des réseaux de neurones en vieillissant causée par une distorsion généralisée. Cette perte d'efficacité neurale obligerait le cortex visuel à utiliser un réseau de neurones plus étendu pour effectuer une même tâche. Donc pour une tâche relativement simple,

un réseau de neurones plus étendu pourrait être sollicité, altérant peu la performance globale. Par contre, pour une tâche plus complexe qui requiert déjà un réseau de neurones étendu, une saturation des ressources causerait des déficits plus importants.

2.2 Augmentation de l'activité neurale

Schmolesky, Wang, Pu, & Leventhal (2000) ont étudié le comportement de neurones du cortex visuel primaire chez le singe à l'aide d'enregistrements extracellulaires. Pour chaque neurone étudié, ils ont enregistré sa vitesse de décharges en fonction d'un stimulus. Les stimuli étaient des images de premier ordre, c'est-à-dire qu'ils étaient définis par la luminance. Ils étaient soit statique pour la condition d'orientation ou dynamique pour la condition de direction. Leurs résultats démontrent que la vitesse maximale de décharges reste la même en vieillissant comparativement à la vitesse minimale de décharges qui augmente. Par exemple, un neurone d'un jeune singe émettra un nombre élevé de décharges pour une orientation spécifique du stimulus et peu de décharges pour une orientation perpendiculaire. Avec le vieillissement, ce même neurone réagira autant à son orientation préférée mais augmentera son nombre de décharges pour d'autres orientations « non préférées » du stimulus. L'activité neurale du cortex visuel primaire augmente donc en vieillissant, ce qui suggère une diminution de l'inhibition. Ainsi, lorsqu'un neurone est grandement activé par son orientation ou sa direction préférée, les autres neurones sélectifs à d'autres orientations sont

moins inhibés causant une distorsion dans le réseau et rendant la tâche de discrimination plus difficile à effectuer.

3 BASE NEURALE DU VIEILLISSEMENT

Pendant plusieurs années, la croyance était que le cerveau, incluant le cortex visuel, perdait des neurones en vieillissant et que ces pertes de neurones causaient des déficits cognitifs et perceptifs. Certaines études rapportaient jusqu'à 50 % de perte neurale (Brody, 1970). Des études plus récentes, utilisant des méthodes différentes pour compter le nombre de neurones, ne rapportent aucune perte de neurones avec le vieillissement pour la plupart des aires cérébrales incluant le cortex visuel (Leuba & Kraftsik, 1994). Ce nouveau constat a ouvert la porte à un débat : quelle est la cause des déficits cognitifs et perceptifs des tâches effectuées par les aires corticales qui ne sont pas affectées par une perte neurale ? Il doit donc y avoir des changements neurobiologiques pour expliquer ces pertes. La connaissance des changements neurobiologiques du cortex visuel qui surviennent avec le vieillissement aide à élaborer certaines hypothèses sur les causes neurales des déficits perceptifs. Cependant, plusieurs changements neurobiologiques se produisent simultanément (Spear, 1993). Il est donc possible qu'un changement neurobiologique corrèle avec une perte perceptive sans pour autant être la cause de cette perte. La présente section, passe en revue trois principales altérations

neurobiologiques : la perte de synapses, la dégénérescence de l'inhibition intra-corticale et la dégradation de la gaine de myéline.

3.1 Perte de synapses

Plusieurs études rapportent une diminution du nombre de synapses avec le vieillissement (Peters, 2002). Il semble y avoir consensus sur un point : une perte d'environ 50 % de synapses se produirait au niveau de la première couche du cortex, ce qui expliquerait son amincissement au profit de la membrane gliale (Peters & Sethares, 2002). Par contre, il y a divergence de points de vue à propos de la perte de synapses pour d'autres couches du cortex. En effet, plusieurs études ne révèlent aucune perte significative de synapses au niveau de d'autres couches (Cragg, 1975; Hutterlocher, 1979), alors que d'autres rapportent une perte d'environ 20 % (Gibson, 1983).

Poe, Linville, & Brunso-Bechtold (2001) ont cherché à savoir quel type de synapses, excitateur ou inhibiteur, était le plus affecté. Ils ont compté le nombre de synapses chez de jeunes singes (moyenne d'âge 23 ans) et de vieux singes (moyenne d'âge 33 ans). Ils ont rapporté que chez les jeunes singes, il y a 4,2 fois plus de synapses excitatrices qu'inhibitrices et que chez les vieux singes, ce ratio augmente à 4,9. Selon ces résultats, ils semblent donc que les synapses inhibitrices soient plus affectées avec le vieillissement.

3.2 Dégénérescence de l'inhibition intra-corticale

L'hypothèse de la dégénérescence de l'inhibition intra-corticale avec le vieillissement va de pair avec les résultats obtenus par Schmolesky et al. (2000) sur l'augmentation de l'activation neurale pour les directions et orientations non préférées. Comme mentionné dans la section précédente, cette perte de discrimination d'orientation et de direction pourrait être causée par une diminution d'inhibition intra-corticale.

Une recherche récente appuie fortement la thèse de la diminution de l'inhibition intra-corticale pour expliquer le déclin perceptif chez les personnes âgées (Leventhal, Wang, Pu, Zhou, & Ma, 2003). Cette étude utilise également la méthode d'enregistrement extracellulaire pour évaluer les comportements des neurones du cortex visuel primaire en fonction de l'orientation et de la direction de mouvement. Les résultats ressemblent à ceux de Schmolesky et al. (2000) pour les jeunes singes et ceux âgés. Ils ont ensuite électrophorétiquement appliqué le neurotransmetteur GABA, un agoniste (muscimol) ou un antagoniste (bucuculline) sur des neurones du cortex visuel primaire et enregistré ce que ces drogues avaient comme effet sur le comportement du neurone. Ils ont trouvé que l'application, chez les singes âgés de GABA et de son agoniste, tous les deux inhibiteurs, aide les neurones du cortex visuel primaire sensibles à l'orientation et à la direction de mouvement à se comporter comme les neurones des jeunes singes. En effet, ceci améliore la discrimination d'orientation et de direction chez les singes âgés en inhibant davantage la réponse aux orientations ou aux directions non préférées. Par

contre, l'application d'une drogue antagoniste au GABA provoque l'effet contraire puisqu'elle diminue la capacité des neurones à discriminer une orientation ou une direction en augmentant le taux d'activation pour les orientations et les directions non préférées. En augmentant l'inhibition intra-corticale, Leventhal et al. (2003) ont donc réussi à redonner aux neurones âgés un comportement similaire à ceux plus jeunes, c'est-à-dire améliorer leur sélectivité à l'orientation et à la direction de mouvement.

3.3 Dégradation de la gaine de myéline

3.3.1 Groupement par synchronisation

Une hypothèse controversée dans la littérature est celle du groupement par synchronisation de l'activité neurale. Le phénomène de synchronisation en soi n'est pas controversé puisqu'il est facilement observable ; des études à l'aide d'encéphalogrammes ont démontré que lors de tâches cognitives, certains groupes de neurones sont activés simultanément (Singer, Engel, Kreiter, & Munk, 1995). Ce qui soulève une controverse, c'est l'hypothèse que le groupement de différents attributs tels la couleur, le mouvement et la forme d'un objet se fasse à l'aide de la synchronisation.

Cette problématique survient du fait que le cortex visuel semble utiliser une approche modulaire pour traiter les différents attributs d'un objet. Ces derniers

seraient donc analysés séparément par des structures différentes. Chaque structure du cortex visuel semble donc spécialisée pour certains attributs bien spécifiques. L'aire V4 semble responsable du traitement des couleurs alors que V5 serait surtout impliquée dans la perception du mouvement. Le problème du groupement est d'unifier ces différents attributs sous un seul et même objet. Par exemple, si un sujet perçoit simultanément deux objets en mouvement : une balle et un bâton de base-ball. Chaque objet possède ses propres attributs pour le mouvement, la couleur, la forme, etc. Le système perceptif doit pairer chaque attribut au bon objet alors que les attributs des deux objets sont perçus simultanément. La partie du cortex visuel qui semble impliquée dans la perception des couleurs, V4, perçoit donc au minimum deux couleurs simultanément. De plus, l'aire présumée impliquée dans la perception des mouvements, V5, perçoit également deux mouvements distincts simultanément. Selon l'approche modulaire, ce fonctionnement cause un nouveau problème : unifier chaque attribut au bon objet, c'est-à-dire, attribuer le blanc à la balle et le brun au bâton sans mélanger les attributs traités simultanément.

La littérature actuelle propose deux hypothèses pour le groupement : la convergence hiérarchique et la synchronisation. Dans la convergence hiérarchique, un système neural se chargerait du problème de groupement (Watt & Phillips, 2000). Dans la synchronisation, les différents attributs d'un même objet seraient groupés par une activation simultanée. Puisque ces deux hypothèses ne se contredisent pas et que beaucoup de preuves supportent les deux (Phillips &

Silverstein, 2003), certains auteurs affirment qu'il est possible et même probable que les deux soient à l'origine du groupement (Treisman, 1999).

3.3.2 Dégradation de la gaine de myéline et synchronisation

Certaines études affirment qu'il est possible que la synchronisation en groupant les attributs soit à l'origine de la conscience sensorielle et perceptive (Engel & Singer, 2001). Suivant cette logique, si la synchronisation est affectée, la conscience sensorielle et perceptive peut également être affectée. Les pertes sensorielles et perceptives constatées lors du vieillissement pourraient donc être causées par un déficit au niveau de la synchronisation. Or, un phénomène important qui affecte l'aspect temporel de l'activité neurale est la gaine de myéline des axones. Cette gaine de myéline augmente la vitesse de transmission le long des axones. Il est maintenant bien connu que cette gaine est affectée de façon non uniforme avec le vieillissement (Nielsen & Peters, 2003) causant un ralentissement inégal de la vitesse de transmission des axones. Cet aspect pourrait donc être à l'origine d'une désynchronisation de l'activation neurale causant certaines pertes perceptives.

4 LES RÉSEAUX DE NEURONES ARTIFICIELS

4.1 Le neurone artificiel

Bien qu'il soit possible de modéliser le comportement d'un seul neurone, généralement le neurone artificiel, aussi appelé unité ou nœud, correspond à une population de neurones similairement configurés. Cette façon de modéliser a l'avantage d'être plus simple à implémenter et permet d'effectuer des tâches plus complexes.

Le réseau de neurones artificiels est composé d'unités reliées entre eux par des poids. Ces poids correspondent à la force synaptique les reliant. La somme de l'activation des unités pré-synaptiques multipliée par leur poids représente le degré d'excitation de l'unité post-synaptique. Le degré d'activation d'une unité est fonction de son degré d'excitation. Puisque l'unité représente un ensemble de neurones biologiques, la fonction d'activation d'une unité est généralement un sigmoïde (O'Reilly & Munkata, 2000).

4.2 Le réseau de neurones artificiels

Il existe plusieurs types d'architecture de réseau, mais le plus utilisé est sans aucun doute les réseaux *feedforwards*. Contrairement aux réseaux récurrents, les unités de ces réseaux affectent seulement ceux de la (ou des) couche(s) suivante(s) sans jamais affecter ceux d'une couche précédente. Dans de tels réseaux, il n'y a donc pas d'activation descendante, mais seulement ascendante. Bien que moins biologiquement plausible, ces réseaux sont plus faciles à implémenter et l'apprentissage est également plus rapide.

4.2.1 Unités d'entrées

Les unités d'entrées n'ont pas d'unités pré-synaptiques. Leur niveau d'activation correspond plutôt au stimulus extérieur. Par exemple, si le stimulus est une image achromatique, chaque unité d'entrée représente la luminosité d'un pixel de cette image (généralement, 0=noir et 1=blanc).

4.2.2 Unités de sorties

L'activation des unités de sorties correspond à la réponse que le réseau donne en fonction du stimulus présenté par l'intermédiaire des unités d'entrée. Dans les simulations effectuées dans ce présent mémoire, la tâche que le réseau doit accomplir est une classification. Le stimulus d'entrée est une image formée de barres d'orientation verticale ou horizontale représentant le premier ou le deuxième ordre. La tâche du réseau de neurones consiste à déterminer l'orientation des barres. Il y a donc deux unités de sorties modélisant deux populations de neurones répondant à deux orientations perpendiculaires. Si l'activation de la première unité de sortie est plus grande que celle de la seconde, le réseau répondra que l'image présentée était d'une certaine orientation et vice versa.

4.2.3 Unités cachées

Les unités cachées sont les plus importantes et surtout les plus intéressantes. Ce sont elles qui effectuent le traitement de l'information entre les unités d'entrées et celles de sorties. Ils sont donc les seules à avoir à la fois des connexions (poids) pré- et post-synaptiques.

4.3 Apprentissage

Bien que les ordinateurs deviennent de plus en plus performants, l'apprentissage peut quand même s'avérer long spécialement pour les gros réseaux effectuant des tâches complexes. Cette section présente les grands processus d'apprentissage des réseaux de neurones artificiels.

4.3.1 Apprentissage hebbian

La forme d'apprentissage biologique la plus connue est l'apprentissage hebbian. Cette méthode stipule que lorsque le neurone post-synaptique et celui pré-synaptique sont activés en même temps, la connexion entre ces deux neurones est renforcée (*long term potential*). Par contre, lorsque le neurone post-synaptique est actif mais que celui pré-synaptique est inactif, la connexion est diminuée (*long term depression*). Si le neurone post-synaptique n'est pas actif, aucun changement se produit au niveau de la connexion (O'Reilly et al., 2000).

Cette forme d'apprentissage est peut-être biologiquement plausible, mais elle est peu efficace pour des tâches de classifications puisqu'elle se contente de faire une corrélation entre deux couches d'unités.

4.3.2 Rétropropagation (*Backpropagation*)

Rumelhart et McClelland (1986) ont introduit un autre algorithme d'apprentissage qui est largement plus efficace pour un apprentissage supervisé : le *backpropagation*. Cet algorithme corrige tous les poids d'un réseau en fonction d'une erreur dans le but de réduire celle-ci. Cette erreur représente une distance entre l'activation des neurones de sorties et leur activation désirée. L'apprentissage peut se résumer en quatre étapes récurrentes : 1) présenter un stimulus ce qui permet de déterminer l'activation des unités d'entrées, 2) calculer la propagation du signal i.e. calculer l'activation des couches cachées puis de la couche de sortie, 3) calculer la distance entre l'activation des unités de sorties et leur activation désirée, c'est-à-dire l'erreur et 4) corriger les poids de façon à réduire l'erreur.

Pour implémenter un apprentissage supervisé, il est nécessaire de connaître l'activation désirée des unités de sorties en fonction du stimulus présenté aux unités d'entrées. Pour chaque poids, la dérivée de la fonction d'erreur par rapport à ce poids est multipliée par une constante nommée taux d'apprentissage. Ensuite, ce produit est soustrait au poids. Pour un taux d'apprentissage infiniment petit, l'erreur diminue nécessairement à chaque correction. Ainsi, l'algorithme converge

assurément vers une erreur minimum. Un taux d'apprentissage plus élevé accélère l'apprentissage puisqu'il permet de plus amples corrections. Cependant, si le taux d'apprentissage est trop élevé, les corrections appliquées seront trop importantes. À ce moment, le nouveau poids se retrouve plus loin du minimum que sa valeur précédente et le réseau n'apprendra jamais. Pour que le réseau ait une bonne vitesse d'apprentissage, le taux d'apprentissage s'avère donc critique.

Depuis 1986, beaucoup d'algorithmes basés sur le *backpropagation* ont été développés dans le but d'accélérer l'apprentissage en facilitant l'ajustement du taux d'apprentissage. Il est possible de les classer en deux catégories : ceux qui ont une approche globale et ceux qui ont une approche locale.

4.3.2.1 Algorithme avec une approche globale

Les algorithmes avec une approche globale se basent généralement sur la fonction d'erreur. Si l'erreur augmente, ceci indique que le taux d'apprentissage est trop élevé et qu'il faut donc le diminuer pour permettre au réseau d'apprendre. Si l'erreur diminue ou reste relativement stable, le taux d'apprentissage est augmenté pour accélérer la vitesse d'apprentissage.

L'approche globale n'est pas la plus habituelle puisqu'elle a un important désavantage : tous les poids du réseau ont le même taux d'apprentissage. Pour cette raison, l'approche locale est généralement plus efficace.

4.3.2.2 Algorithme avec une approche locale

Avec l'approche locale (delta-bar-delta : Jacobs, 1988; RPROP : Reidmiller & Heinrich, 1993), chaque poids a son propre taux d'apprentissage. Cette approche utilise le signe de la correction du poids i.e. la dérivée de ce poids pour ajuster son taux d'apprentissage. Si le poids subit une correction toujours dans la même direction, son taux d'apprentissage est augmenté pour accélérer l'apprentissage. Par contre, si la correction change continuellement de direction, c'est que le poids oscille autour d'un minimum. Ceci indique que ces corrections sont trop grandes puisqu'il chevauche continuellement de part et d'autre ce minimum. Le taux d'apprentissage de ce poids est alors diminué de façon à permettre au poids de converger vers son erreur minimum.

5 MODÉLISATION D'UN RÉSEAU DE NEURONES BIOLOGIQUES

5.1 Modélisation d'un seul neurone

Il est possible de développer un modèle dans lequel un neurone est modélisé à l'aide d'un grand nombre de variables. Il est également possible de lier de tels neurones entre eux pour simuler un réseau de neurones. Ce type de modélisation a l'avantage de modéliser différentes parties d'un neurone de façon très précise, mais

il implique deux désavantages importants. Premièrement, il demande beaucoup de calcul pour un seul neurone. Deuxièmement, il est très difficile d'apprendre à de tels réseaux à effectuer des tâches complexes.

En effet, les mécanismes de l'apprentissage neural chez l'humain sont peu connus. Paradoxalement, il est plus facile d'apprendre une tâche complexe à un réseau de neurones artificiels lorsque la complexité des neurones artificiels est largement simplifiée puisque ceci accélère la vitesse de calcul. Comme vu précédemment, pour le présent mémoire, il n'y a pas de modélisation pour un neurone, mais plutôt pour un ensemble de neurones similairement configurés, nommé unité.

5.2 Simulation

Les deux éléments fondamentaux du réseau de neurones sont les unités et les poids. Chacun de ces éléments peut être inhibiteur ou excitateur. Comme un neurone excitateur ne possède, généralement, que des synapses efférentes excitatrices et qu'un neurone inhibiteur ne possède, généralement, que des synapses efférentes inhibitrices, une unité ne possède que des poids efférents du même type. Une unité avec des poids efférents négatifs, c'est-à-dire inhibiteurs, est dite inhibitrice et une unité avec des poids efférents positifs, c'est-à-dire excitateurs, est dite excitatrice. Un tel réseau de neurones artificiels est donc analogue à un réseau de neurones biologiques : il peut effectuer des tâches complexes à l'aide d'éléments simples (neurones) reliés entre eux (synapses).

Le vieillissement peut être modélisé en introduisant une distorsion au niveau des poids ou au niveau de l'activation des unités. Cette distorsion peut s'appliquer seulement aux éléments excitateurs, seulement aux éléments inhibiteurs ou aux deux. Puisqu'on suppose que les pertes cognitives et perceptives sont causées par un changement neural généralisé, la distorsion peut consister soit à une augmentation uniforme de tous les éléments (unité ou poids), soit à une diminution uniforme de tous les éléments ou soit à l'introduction de bruit i.e. que certains éléments sont augmentés et d'autres diminués. En résumé, toutes ces altérations possibles constituent au total 18 simulations ($2 \times 3 \times 3$) : distorsion au niveau des deux éléments, les unités ou les poids (2), distorsion des éléments excitateurs, inhibiteurs ou les deux (3) et distorsion de type augmentation uniforme, diminution uniforme ou introduction de bruit (3). De plus, à ces 18 possibilités, trois autres simulations s'additionnent pour simuler la perte de synapses excitatrices, inhibitrices ou, excitatrices et inhibitrices.

5.3 Modélisation des hypothèses biologiques

La perte de synapse est évidemment modélisée par la perte de poids (connexions). La dégénérescence de l'inhibition intra-corticale est modélisée par une diminution uniforme des poids inhibiteurs puisqu'elle est responsable des connections inhibitrices du cortex. La dégradation de la gaine de myéline est par contre moins directement modélisé car il n'y a pas de facteur temps dans le réseau de neurones

artificiels utilisé dans ce mémoire. Il est donc difficile de simuler un problème au niveau de la synchronisation. Par contre, en supposant qu'une désynchronisation causerait du bruit au signal d'une population de neurones similairement configurés, il est possible de simuler indirectement la dégradation de la gaine de myéline par du bruit au niveau de l'activation des unités.

CONTEXTE DE L'ÉTUDE

Comme vu précédemment, plusieurs ANBs se produisent au niveau des neurones lors du vieillissement. Ces ANBs sont à l'origine de déficits subtils qui peuvent être détectés chez les aînés. L'hypothèse prédominante dans la littérature est que plus une tâche est complexe, plus le déficit sera prononcé.

Le présent mémoire vise à mieux déterminer les causes neurales des pertes perceptives occasionnées par le vieillissement. Pour ce faire, des simulations artificielles d'ANBs pouvant expliquer les pertes perceptives dues au vieillissement ont été réalisées sur un RNA effectuant une tâche simple et une tâche complexe. Le rapport des déficits dus aux ANBs pour ces deux tâches comparativement à celui observé chez l'humain dû au vieillissement a servi de critère pour déterminer la plausibilité des différentes hypothèses concernant les ANBs.

Lors des recherches, une difficulté s'est présentée : l'apprentissage de la tâche complexe par le RNA. Pour permettre au RNA d'apprendre cette tâche, un nouvel algorithme a dû être développé. Le premier article est consacré à cette fin, tandis que le second porte sur les simulations d'ANBs.

Chapitre II

ARTICLE 1

The generalized no-decrease adaptive method
for large-slow-learning neural networks

(article soumis à *Neural Networks*)

Rémy Allard, Jocelyn Faubert, Ph.D.

ABSTRACT

In the present study, we introduced a complex task (texture task) where standard backpropagation and its variants failed to produce a solution within a reasonable time. The generalized no-decreased adaptive method developed in this paper was motivated by this failure. Instead of using the weight derivative sign to adjust the learning rate, this method uses a trial and error heuristic where two identical networks with different learning rates are computed in parallel. Their performances (error rates) are used as the criterion to adjust the learning rate. Applying this method for the texture task showed that the learning rate was not only automatically adjusted, but also that the learning speed was greater than the optimal fixed learning rate. We showed that our generalized no-decrease adaptive method for large-slow-learning neural networks is useful for this complex learning task. Future experiments should assess its robustness to other complex tasks.

INTRODUCTION

Backpropagation (BP) (Rumelhart, Hinton & Williams, 1986) is probably the most commonly used learning rule for artificial neural networks. For each iteration, the weights are updated by the following rule:

$$\Delta w_i = -\eta \frac{\partial J}{\partial w_i}$$

where w_i is the i^{th} weight in the network, J the error function and η the learning rate (LR), which is a positive constant. The objective of BP is to minimize the error function J . The partial derivative of that function according to i^{th} weight with the minus sign ($-\partial J/\partial w_i$) indicates the direction and the relative size of the change. The LR (η) globally scales the size of changes in the entire network. With standard BP, all the weights in the network have the same LR. To achieve good results using this algorithm η must be properly tuned. Furthermore, η is problem dependent. This explains why BP may sometimes be difficult to use.

If η is too high, weight changes will be too big causing the new weight values to be further than they were from their optimal values and the algorithm will never converge (the network will saturate). If η is too small, learning will be very slow.

Since there is no mathematical rule that determines the optimal LR, the typical way to adjust the LR is by trial and error. Typically, we start by using the value 0.1. If the network saturates, we decrease the LR. If convergence is too slow, we increase it. Adjusting the LR when the network is small and learns relatively

quickly is not difficult. However, when a large network has to learn a complicated task, finding a good LR can be harder. This is especially true when the network needs many iterations before its performance gets better than the optimal a priori error rate (OAPER). We define the OAPER as the minimized error rate when nothing is known about the input pattern. For example, in a classification task with a single output unit that has a target value of either -1 or $+1$ equally distributed and using a quadratic error function, the OAPER would be equal to the error rate when the output unit fires 0.

Many adaptive methods (AMs) for the LR already exist in the literature; however, most have two or more parameters to adjust and one of those parameters directly influences the LR. This parameter may be just as difficult to tune as the LR using standard BP. For example, Delta-bar-delta (Jacobs, 1988) still has a parameter to tune: an incremental constant that has a direct influence on the LR. Most AMs often increase the learning speed once their parameters have been properly tuned. However, it has been argued that they may not be worth implementing because of the time required to tune those parameters and because their performance may not be significantly improved (Alpsan, Towsey, Ozdamar, Tsoi, & Ghista, 1995).

THE TEXTURE TASK

In this section, we describe a task where the standard BP algorithm and its variants failed to produce a solution within a reasonable time. The AMs developed in the present study were motivated by this failure.

We wanted to create a network able to discriminate the orientation (horizontal or vertical) of first and second order stimuli (images), similar to the static task the subjects had to perform in a human psychophysical experiment by Habak and Fauber (2000). Figure 1 shows similar examples of those images. First order stimuli are gratings defined by luminance with noise added and second order stimuli are gratings defined by contrast. Simulation showed that discriminating the orientation of first order stimuli was relatively easy. However, making a network learn the same task for second order stimuli was much more complicated. We found that the network needed more units and that the LR was difficult to tune. The main reason why tuning was difficult was that the network took many iterations before its error rate dropped below the OAPER, and this, even with the optimal LR (see first experiment below).

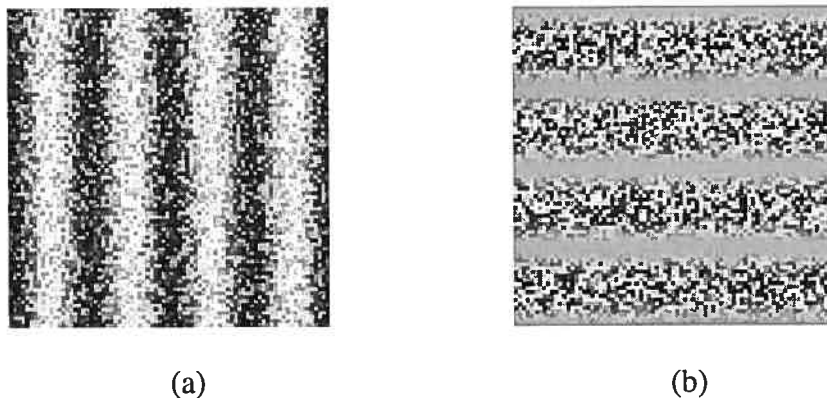


Figure 1. (a) is a first order stimulus (a grating defined by luminance with added noise) and (b) is a second order stimuli (a grating defined by contrast).

These simulations were accomplished with 8-by-8 pixel images. The reason we wanted to create an artificial network able to discriminate the orientation of first and second order images was to compare the thresholds (grating contrast necessary to perceive the orientation) between biological and artificial neural networks. However, the artificial network thresholds were very high due to the small image size. To allow lower thresholds we needed larger images requiring larger networks. Doubling the stimulus size (height and width) implied quadrupling the size of each non-output layers giving the network approximately sixteen times more connections. Therefore, each iteration required sixteen times more computation. Since it took hours before the network showed significant signs of learning (before the performance dropped below the OAPER) with the optimal LR, it could take days or months with larger images (16-by-16 or 32-by-32) before it could show any significant signs of learning and this, once the LR is properly tuned. Therefore, adjusting the LR for larger images revealed to be impractical for

the texture task (discriminating the orientation of second order stimuli). Other AMs such as Delta-Bar-Delta (Jacobs, 1988), SuperSAB (Tollenaere, 1990) and RPROP (Reidmiller & Heinrich, 1993) were tried but no significant improvement was noted.

FIRST EXPERIMENT: AN EXHAUSTIVE SEARCH FOR THE OPTIMAL LEARNING RATE

As discussed above, adjusting the LR for large-slow-learning neural networks is difficult. This is due to the fact that the network needs a long computation time before it is possible to know if all the network parameters (i.e. LR, number of hidden layers, number of units...) have been properly set. If the network does not have enough units to learn the required task, how can one know why the network cannot learn? Many causes could be guessed. The LR may not be well tuned. It is also possible that the network had seen too few iterations or that there is not enough units. Adjusting different parameters can be difficult when the network does not show any signs of learning after many iterations.

Our first experiment consisted in finding, by an exhaustive search, the optimal fixed LR for the texture task. This experiment had two goals: show how difficult it can be to adjust the LR for certain problems and use those results to compare the performance of AMs developed in further experiments.

First experiment: methodology

As mentioned earlier, the network had to learn to discriminate the orientation of second order stimuli. The range of LRs was chosen so that the lowest LR would show no signs of learning and the highest LR would clearly bring the network to saturation. If we suppose that the error function can be approximated by a quadratic function (according to the weights), then the highest LR to generate convergence should be twice the optimal LR (Duda, Hart, & Stork, 2001). This is why a factor of 2 separated two consecutive LRs. Eighteen different LRs were tested: from 2^{-23} to 2^{-6} (LRs greater than 2^{-6} produced saturation). All networks had the same initial weight values and used the same training set.

The goal of this study was to develop an AM that can find, within a reasonable time, a good LR for large networks that normally take so much time to learn that it is practically impossible to tune the LR. However, before developing this AM, we first needed to find the optimal LR to allow a performance comparison. For this reason, we had to use an image size that allowed a LR tuning within a reasonable time. Therefore, we used 8-by-8 pixel images although the AM developed is actually going to be used for larger ones.

Each pixel of an image corresponded to a unit of the input layer. An image was presented with stripes having two different textures. The task was to identify the

orientation of the stripes (horizontal or vertical). A texture was defined by an amplitude of luminance variation (A). All the textures had the same luminance average (L_{avg}). Mathematically, this means that the luminance value for each pixel of a given texture was randomly chosen between $L_{avg}-A$ and $L_{avg}+A$. For each image, two different textures (amplitudes) were randomly chosen. The stripes were defined by a square wave. Given that we have 8-by-8 pixel images, the square wave of each image could have one of three different periods: 2, 4 or 8 pixels. The square wave was not required to start at the beginning of a cycle; it could be shifted in any possible phase. So the orientation, the two amplitudes, the cycle and the shifted position were all randomly selected for each pattern (image). Figure 2 shows stimuli examples. These stimuli are similar to the second order stimulus of figure 1 with the exceptions that the stripes are defined by a square wave instead of a sinusoid and that the images are smaller.

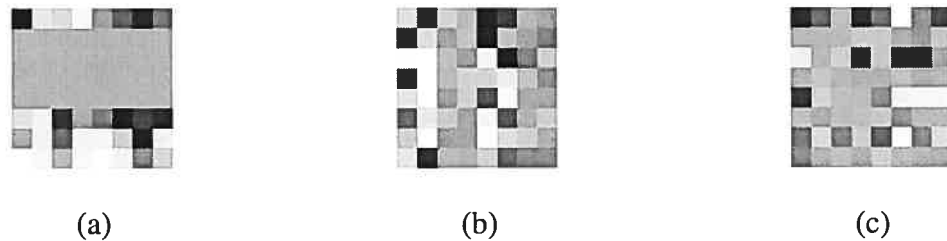


Figure 2. (a) is horizontally oriented, the cycle has a period of 8 and the two amplitudes are 0 and 1. (b) is vertically oriented, the cycle has a period of 4 and the two amplitudes are 0.25 and 1. (c) is horizontally oriented, the cycle has a period of 2 the two amplitudes are 0.2 and 0.9.

Even if the LR is probably the most difficult parameter to tune, there are many other parameters to adjust in a network. Bellow, we simply enumerate the different parameter settings. We tried to choose each parameter in their most standard settings.

Input Layer. We scaled the activation of the input units between -1 and $+1$. This means that L_{avg} was set to 0 and A (the amplitude) for each texture was between 0 and 1.

Output Layer. As a typical categorical task, there were two output units with target values of either $[-1, +1]$ or $[+1, -1]$ depending on the orientation of the stripes.

Activation function. The activation function that was used is the standard hyperbolic tangent (Duda et al., 2001):

$$f(net) = a \cdot \tanh(b \cdot net)$$

where a equals 1.716 and b equals $2/3$.

Number of hidden units and layers. Since our training set was infinite (each pattern was presented only once), overfitting was impossible. For this reason, we preferred having too many units rather than too little. Two hidden layers of 128 units (twice as much as the input layer) were used. A bias unit was also connected to each non-input unit.

Initializing weights. The weights were initialized between $-3/\sqrt{f}$ and $3/\sqrt{f}$, where 'f' is the fan-in to the unit. As mentioned above, each network was always initialized with exactly the same weight values.

Error function. The error function used was the sum of the square of the difference between the output and the desired output (target):

$$J(\mathbf{w}) = \frac{\|\mathbf{t} - \mathbf{z}\|^2}{2}$$

where \mathbf{w} represents all the weights in the network, \mathbf{t} the target vector and \mathbf{z} the activation vector of the output units.

Stopped training criterion. The training was stopped after one hundred million iterations.

First experiment: results and discussion

Figure 3 shows the evolution of the error rate as a function of the LR. Given that the OAPER was 1 (when both output units fire 0), an averaged error rate greater than 1 implied that the network did not show signs of significant learning. LRs between 2^{-12} and 2^{-6} did not make the network either saturate or significantly learn. The network showed signs of different learning speeds when the LR was equal or less than 2^{-13} (greater the LR gave greater learning speed). The first LR that made

the error rate fall below the OAPER was 2^{-13} , followed by 2^{-14} , 2^{-15} and 2^{-16} . Given these results, we can reasonably assume that the LRs below 2^{-16} would eventually learn with more iterations even though they did not yet perform better than the OAPER.

Figure 4 shows the error rate for each LR after 20, 50 and 100 million iterations. As we can see, no LR caused an error rate smaller than the OAPER within 20 million iterations.

Figure 5 demonstrates the performance of the optimal fixed LR. This performance was later used as a base line to evaluate the performance of AMs developed in the present study. It is important to note that this performance was achieved after an exhaustive search, which would be impractical using larger images because of the computation time required.

This first experiment clearly demonstrates how difficult it can be to adjust the LR. Indeed, finding a good LR by the trial and error approach can be a very long process for a large-slow-learning network. With too few iterations (less than 20 millions in our case), an optimal LR may even be considered invalid since it would not show signs of significant learning.

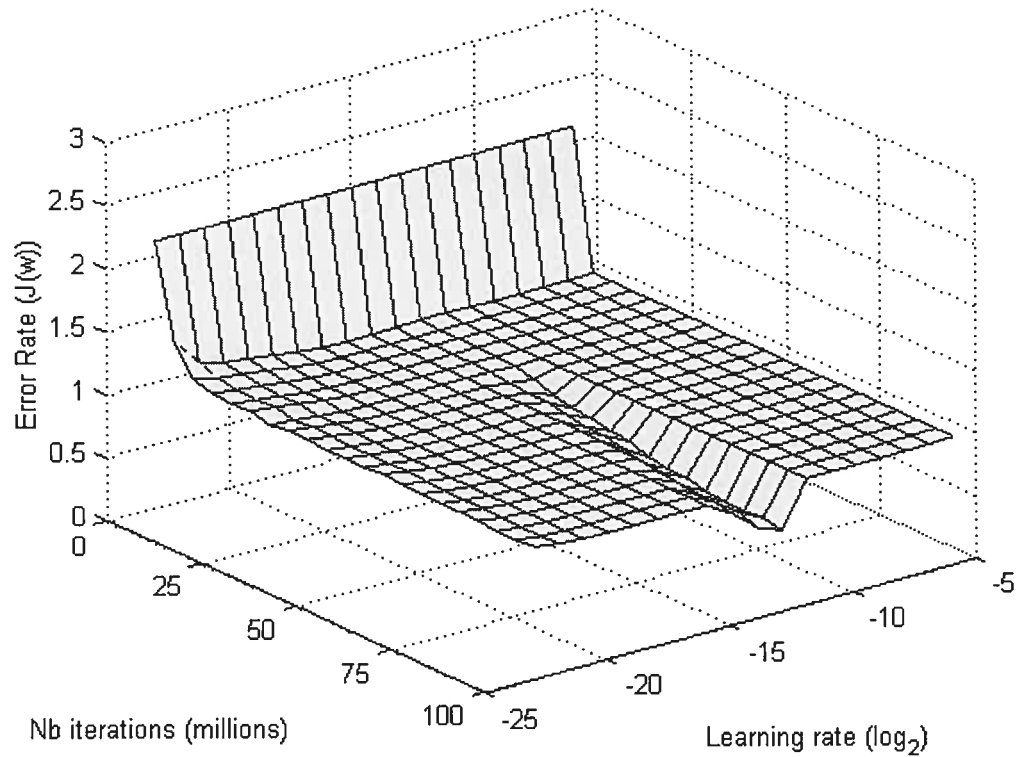


Figure 3. Learning performance of eighteen different networks each of them having their own fixed LRs (2^{-23} to 2^{-6}) on the texture task. The error rate is the error produced at each iteration during the learning. To smooth the curve, we apply a kernel convolution to the vector of the one hundred million iterations of each LR with the vector $(1, 2, 3, \dots, 99\,999, 100\,000, 99\,999, \dots, 3, 2, 1)^t$. The initial error rate was set to 2.1 since it was the averaged error rate produced by the initial weights.

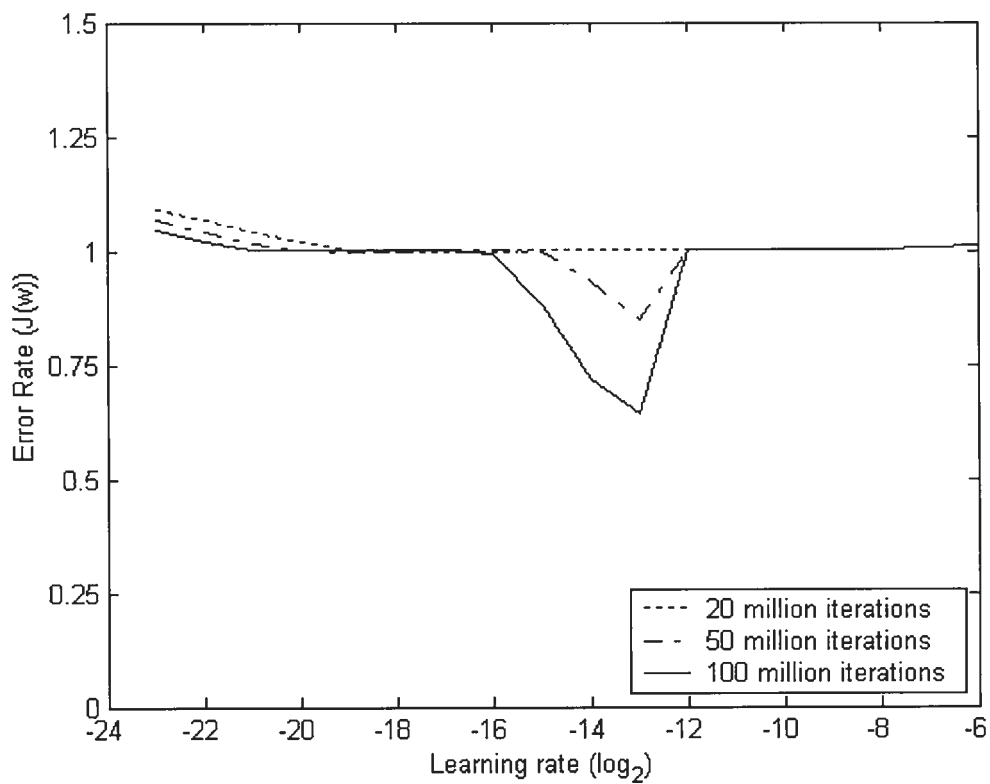


Figure 4. Error rates of different LR's at different learning stages. The error rate was averaged with a kernel convolution as in the preceding figure.

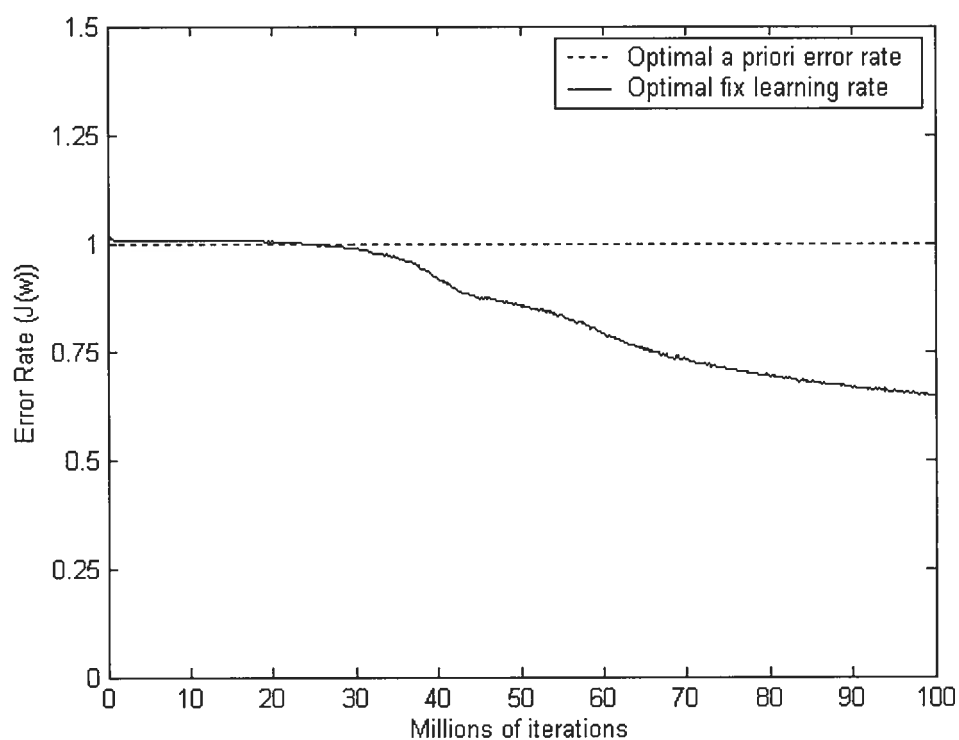


Figure 5. Evolution of the error rate for the optimal fixed LR. A kernel convolution was applied to the error rate as described in figure 3.

SECOND EXPERIMENT: THE NAÏVE ADAPTIVE METHOD (NAM)

Since we cannot know the optimal LR a priori, the conventional heuristic for adjusting the LR is by trial and error. The NAM developed in this experiment is inspired by this simple and intuitive heuristic. The approach of this NAM is defined in four looping steps. First, we compute the learning of two identical weight sets with different LR for a given number of iterations. Then we observe which weight set demonstrated the lowest error rate and discard the other. The third step consists in increasing or decreasing the LR according to the weight set kept. Finally, we start over to the first step with two weight sets identical to the weight set kept and with the new LR. Specifically, we apply the BP algorithm to two identical weight sets ($\mathbf{w}_{low\ t}$ and $\mathbf{w}_{high\ t}$, representing the two weight sets at time t) having different LR ($\eta_{low\ t}$ and $\eta_{high\ t}$ respectively, where $\eta_{high\ t} = \eta_{low\ t} \cdot K$) for N iterations. K is a constant greater than 1 that separates the two LR. For the same reason as we used a factor of 2 separating the LR in the preceding experiment, we recommend setting K to 2. After N iterations, if the weight set that demonstrated the lowest error rate was $\mathbf{w}_{low\ t}$, we set $\mathbf{w}_{low\ t+1} = \mathbf{w}_{high\ t+1} = \mathbf{w}_{low\ t}$, $\eta_{low\ t+1} = \eta_{low\ t} / K$ and $\eta_{high\ t+1} = \eta_{low\ t}$. Otherwise, we set $\mathbf{w}_{low\ t+1} = \mathbf{w}_{high\ t+1} = \mathbf{w}_{high\ t}$, $\eta_{low\ t+1} = \eta_{high\ t}$ and $\eta_{high\ t+1} = \eta_{high\ t} \cdot K$. We then follow by comparing $\mathbf{w}_{low\ t+1}$ and $\mathbf{w}_{high\ t+1}$ using $\eta_{low\ t+1}$ and $\eta_{high\ t+1}$ respectively for another N iterations.

To evaluate which weight set demonstrated the lowest error rate, we can simply use the error rate that occurred at each iteration during the learning process.

However, since the two weight sets are initially the same, we want to give less importance to the error rates that occurred at the first iterations compared to the latter ones where the two sets of weights differed more from one another. For this reason, instead of simply averaging the error rates, we used an exponentially weighted average:

$$W(\mathbf{e}) = \frac{\sum_{i=1}^N [i \cdot e_i]}{\sum_{i=1}^N i}$$

where N represents the number of iterations and e_i the error rate at the i^{th} iteration.

An advantage of the NAM is that it eliminates the time spent searching for an optimal (or good) LR. Another advantage is that it allows the LR to vary over time. This second advantage could imply a better performance than the optimal fixed LR since the optimal LR may not be the same at different learning stages. For example, a large LR could generate a faster solution, and once this good solution is reached, a small LR may permit a better convergence.

Given that the AM works well, the only important disadvantage compared to the standard BP would be that two weight sets are used. This implies that each iteration would be two times slower.

Second experiment: methodology

The parameters of the network were the same as the ones in the first experiment. The weights were initialized with the same values and we used the same training set. The only difference with this methodology is that instead of trying each LR one by one, we used the NAM to adjust the LR. Since we prefer a slow convergence to saturation, and that the LR can increase exponentially, we have set the initial LR relatively low: 2^{-25} . 1000 iterations were used to compare two weight sets.

Second experiment: results and discussion

After fifteen thousand iterations the LR was higher than the optimal LR (2^{-11}). This indicates that the initial LR was not critical. However, as the error rate got closer to the OAPER the LR decreased drastically to settle between 2^{-40} and 2^{-65} so the network never learned significantly. This behavior could be explained by the fact that before the network learns significantly, minimizing the error rate consisted in converging to the OAPER. Having a small LR would reduce the error rate by permitting a better convergence to the OAPER but would not accelerate the learning. This could be a reason why the LR became very low using the NAM.

THIRD EXPERIMENT: THE NO-DECREASE ADAPTIVE METHOD (NDAM)

To correct the tendency of the LR to decrease, we modified the NAM by decreasing the LR only if the error rates weighted average of the lowest LR was smaller than the one of the highest LR ($W(e_{low}) < W(e_{high})$) and that the differences between error rate pairs ($W(e_{high} - e_{low})$) was significantly different from zero. Therefore, if $W(e_{low}) < W(e_{high})$ but $W(e_{high} - e_{low})$ was not significantly different from zero the LR was increased instead of being decreased. We hoped this modification would force the LR to increase when it is small without causing saturation since there should be a significant difference between weights that saturate and ones that do not.

The *EWF* test

To test if $W(e_{high} - e_{low})$ was significantly different from zero, we used a modified version of the F test that we called the exponentially weighted F test (*EWF* test). The F test was modified to give more importance to the latest iterations as opposed to the initial ones. Therefore, the modification affected the way to calculate the error of the full and restricted models. For the full model, the mean (μ_f) of the data set (e) was replaced by an exponentially weighted average as described above ($\mu_f = W(e)$). Since the mean of the restricted model (μ_r) does not depend on the data

set, it was not modified (for us $\mu=0$ since we want to test if the difference is significantly different from zero). Usually, the F test error (E_F) is calculated using the sum of square of the differences between \mathbf{e} and μ :

$$E_F(\mathbf{e}, \mu) = \sum_{i=1}^N [(e_i - \mu)^2]$$

Instead, an exponentially weighted sum was used to calculate the EW_F test error (E_{EW_F}):

$$E_{EW_F}(\mathbf{e}, \mu) = \sum_{i=1}^N [i \cdot (e_i - \mu)^2]$$

The critical number of iterations

Before testing this new AM, we can anticipate that the LR will directly depend on the number of iterations chose to compare two different LR's (N) since the probability of having a significant difference ($P[S]$) also depends on N . If N is too low, the difference between the error rates will rarely be significant implying that the LR will frequently increase and may eventually be too high causing saturation. If N is too high, the difference will often be significant and the algorithm will be similar to the NAM: the LR will be very low. So unlike the NAM, the value of N is critical. Consequently, N must be set so that the LR remains relatively stable. To accomplish that, the probability of increasing ($P[I]$) should be equal to probability of decreasing ($P[D]$), that is

$$P[I] = P[D] = 0.5$$

Since

$$P[D] = P[W(e_{\text{low}}) < W(e_{\text{high}})] \cdot P[S]$$

where $P[W(e_{\text{low}}) < W(e_{\text{high}})]$ is the probability that the weighted average of the error rates of the lowest LR is smaller than the one of the highest LR, we want

$$P[S] = 0.5 / P[W(e_{\text{low}}) < W(e_{\text{high}})]$$

The problem is now shifted since $P[W(e_{\text{low}}) < W(e_{\text{high}})]$ is not known and varies over time, which implies that $P[S]$ also varies over time. We approximated $P[W(e_{\text{low}}) < W(e_{\text{high}})]$ at time t (p_t) by the following algorithm:

$$p_{t+1} = \alpha \cdot p_t + (1 - \alpha) \cdot I_t$$

where α is a decayed factor (a positive constant less than 1) and I_t equals 1 if $W(e_{\text{low}}) < W(e_{\text{high}})$ at time t and 0 otherwise. We recommend setting $p_0 = 1$ and $\alpha = 0.9$.

In summary, by knowing $P[W(e_{\text{low}}) < W(e_{\text{high}})]$, the desired $P[S]$ is also known and then N can be set to cause the desired $P[S]$ that will imply that $P[I] = P[D]$. Therefore, we enabled N to vary over time so it remained at the critical value. To do this, we borrowed and modified a technique used in psychophysics to find a threshold: the QUEST (Watson & Pelli, 1983).

The QUEST

In the present section, the QUEST (Watson et al., 1983) will be briefly described. The notation used will not be exactly the same as the one used by Watson and Pelli (1983), however it is mathematically equivalent.

The QUEST has a wide applicability range. Therefore, its syntax is general and we will begin by defining this syntax according to our present problem. In the present study, an “intensity” corresponds to a number of iterations used to compare two LRs. A “trial” is the application of a given intensity, which is the learning of two initially identical weight sets with different LRs for a certain number of iterations (an intensity). A “success” corresponds to a significant difference in the error rates weighted average after a trial (a “failure” corresponds to a non-significant difference). The “threshold criterion” is the probability of a success that defines the threshold, for us $P[S]$ (the desired number of iterations). Finally, the “threshold” is the intensity at which the probability of having a success is $P[S]$.

To use the QUEST we must initially define a threshold criterion (τ) and approximate a psychometric function ($\Psi(x)$), which estimates the probability of a success as a function of the distance from the threshold in log intensity (therefore $\Psi(0) = \tau$).

The other function that must be estimated is the prior probability distribution function (pdf) of the log threshold intensity ($\text{PDF}_0(T)$). Note that this function is only useful for the first few trials; the QUEST will find a posterior pdf afterwards.

After initializing these settings, the QUEST can be used to find the posterior pdf of the threshold with three looping steps (t for the t^{th} loop):

- 1) $I_t = \log$ intensity of the PDF_{t-1} peak
- 2) The t^{th} trial at intensity $\exp(I_t)$
- 3) $\text{PDF}_t(T) = \text{PDF}_{t-1}(T) \cdot \begin{cases} \Psi(I_t - T) & \text{if trial } t \text{ is a success} \\ 1 - \Psi(I_t - T) & \text{if trial } t \text{ is a failure} \end{cases}$

The FQUEST

For our application, two problems occurred using the QUEST. First, this technique is used to find a single threshold that does not vary over time. Second, when using this technique, the threshold criterion (for us $P[S]$) must be set a priori and remains fixed afterwards. Therefore, we modified the QUEST to enable the threshold and the threshold criterion to vary over time. We called this new technique: the flexible QUEST (FQUEST).

To give the pdf the flexibility to readjust itself to a new threshold, we weighted the importance of the successful and failure functions according to the trial. At each

FQUEST iteration, we decayed the weight of the previous pdf. To do so, the third step was replaced:

$$3) \text{PDF}_t(T) = \text{PDF}_{t-1}(T)^\lambda \cdot \begin{cases} \Psi(I_t - T) & \text{if trial } t \text{ is a success} \\ 1 - \Psi(I_t - T) & \text{if trial } t \text{ is a failure} \end{cases}$$

where λ is a positive constant less than 1. This implies that the latest trials affect more the pdf of the threshold than the previous ones. Therefore, if the threshold varies over time, the pdf should be able to readjust itself to the new threshold.

The other problem with the QUEST is that we need to set a priori the desired threshold criterion (for us $P[S]$). However, the threshold criterion can be changed by shifting the psychometric function along the abscissa. If ε_t and τ_t are the shifted position and the threshold criterion at trial t respectively, they should obey the following equation:

$$\Psi(\varepsilon_t) = \tau_t \quad \text{or} \quad \varepsilon_t = \Psi^{-1}(\tau_t)$$

Therefore, the third step can be replaced by the following:

$$3) \text{PDF}_t(T) = \text{PDF}_{t-1}(T)^\lambda \cdot \begin{cases} \Psi\left(I_t - T + \Psi^{-1}(\tau_t)\right) & \text{if trial } t \text{ is a success} \\ 1 - \Psi\left(I_t - T + \Psi^{-1}(\tau_t)\right) & \text{if trial } t \text{ is a failure} \end{cases}$$

This latest modification enables the threshold criterion to be changed over time (for the purpose of the present study: $\tau_t = 0.5/p_t$).

With the FQUEST the fact that the threshold and the threshold criterion varies over time does not cause a problem. The FQUEST can be used to estimate an intensity (a number of iterations) that will approximate a desired probability of success ($P[S]$) that varies over time.

Third experiment: methodology

The prior pdf of the log threshold intensity ($PDF_0(T)$) was initialized as following:

$$PDF_0(T) = \exp\left(-\left((\mu - T)/\sigma\right)^2\right)$$

where μ is the peak of the pdf and σ the standard deviation. μ and σ were both set to 5, which means that the initial number of iterations used to compare two LRs was set to 148. It is important to specify that the initialization of the threshold pdf is not very important since it is decayed at each trial.

Since $\exp(20)$ is greater than one hundred thousand iterations (our stopped training criterion), a possible range of 0 to 20 log intensity units for the threshold pdf was sufficient. Setting the precision to 0.01 log intensity units implied that a two-thousand-index table was used to approximate the threshold pdf.

The psychometric function used was based on the Weibull distribution (Weibull, 1951):

$$\Psi(x) = 1 - \max\left(\delta, (1 - \gamma)\exp\left(-10^{\beta x}\right)\right)$$

where δ truncates the function to avoid a probability near 1 that would result in extremely low values in the failure function (step 3). γ is the probability of success at zero intensity (the probability of having a significant difference between two LR's after a single iteration) and β determines the slope of the function. δ was set to the typical 0.01, γ to 0 and β to 0.35 (3.5/10) so the function varies significantly over 10 log intensity units.

Concerning the third step of the FQUEST, the parameter λ was set to .9.

The decayed factor used to estimate $P[S]$ (α) was set to 0.9 and a confidence level of 95% was used for the *EW*F test.

The other parameters were initialized or set to the same values as the previous experiment.

Third experiment: results and discussion

The error rate converged to the OAPER, meaning that the network had not yet learned. By looking at the behavior of the LR (figure 6) we can see that it remained relatively stable around 2^{-21} . During the first ten thousand iterations the LR increased up to 2^{-10} (does not appear on the graph), before it decreased to 2^{-21} . Given the fact that no LR over 2^{-13} could do the required task and that any LR equal or lower seemed to eventually learn, we can consider that the LR remained stable and would eventually cause learning but was still low using the NDAM.

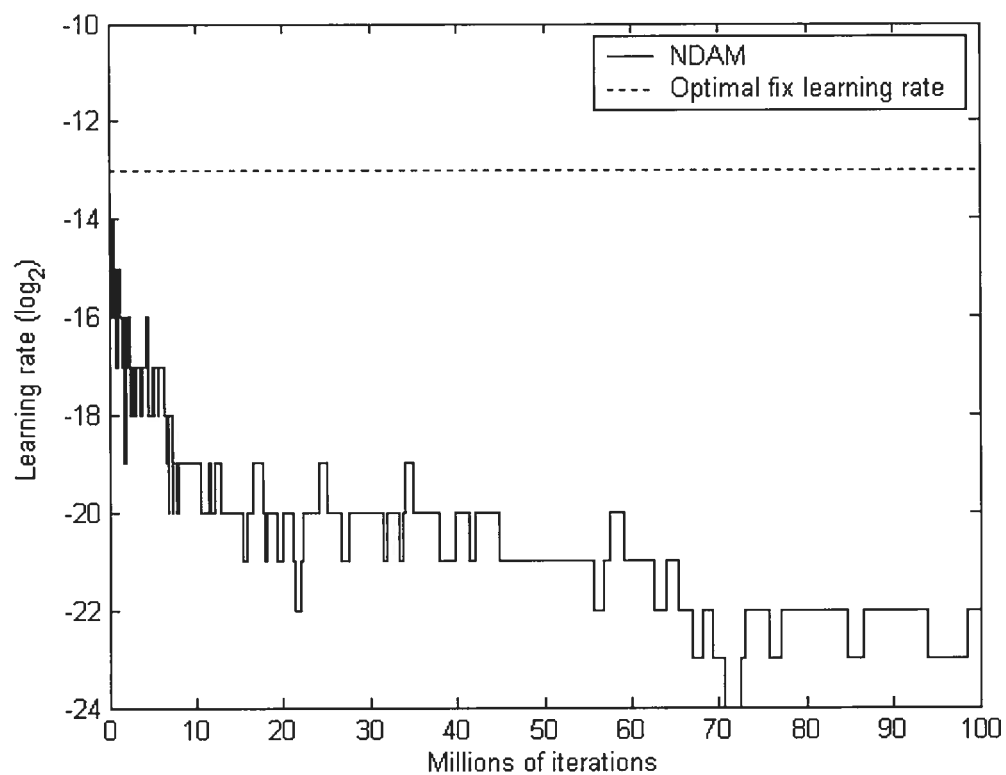


Figure 6. Evolution of the LR throughout the learning process using the NDAM.

FOURTH EXPERIMENT: THE GENERALIZED NDAM (GNDAM)

An important disadvantage of the preceding AM is its global approach; all the weights in the network have the same LR. Applying the NDAM to each weight in the network would be impractical because of the time required to adjust them. However, the weights may be separated in different groups, each of them having their own LR. Then, applying one FQUEST trial of the NDAM to one group at a time, while the other groups maintain a fixed LR may increase performance. Pseudo-code of this last AM is presented in an annex.

Fourth experiment: methodology

An issue in this approach is how we are going to separate the weights in different groups. Separating them in too many groups will result in a slow learning AM given that many iterations will be required to obtain a significant difference between the two LRs. In this last experiment, three groups were defined. The first, second and last group were represented by the input-weights of the first hidden, second hidden and output layer respectively.

The other settings were the same as the previous experiment. The LR of each group was initialized to 2^{-25} and each threshold pdf was initialized the same way as in the previous experiment.

Fourth experiment: results and discussion

Figure 7 shows that the three LRs remained relatively stable throughout the learning process (2^{-9} , 2^{-14} and 2^{-17}) and that the closer the weight group was from the output layer, the smaller the LR was.

As we can see in figure 8, learning began earlier with the GNDAM than with the optimal fixed LR. This new AM took less than two million iterations before the error rate dropped below the OAPER compared to more than twenty million iterations for the optimal fixed LR. Even if we consider that the iterations of the GNDAM needs two times more computation than standard BP (the computation time curve on the graph), this AM was still more than five times faster to decrease below the OAPER.

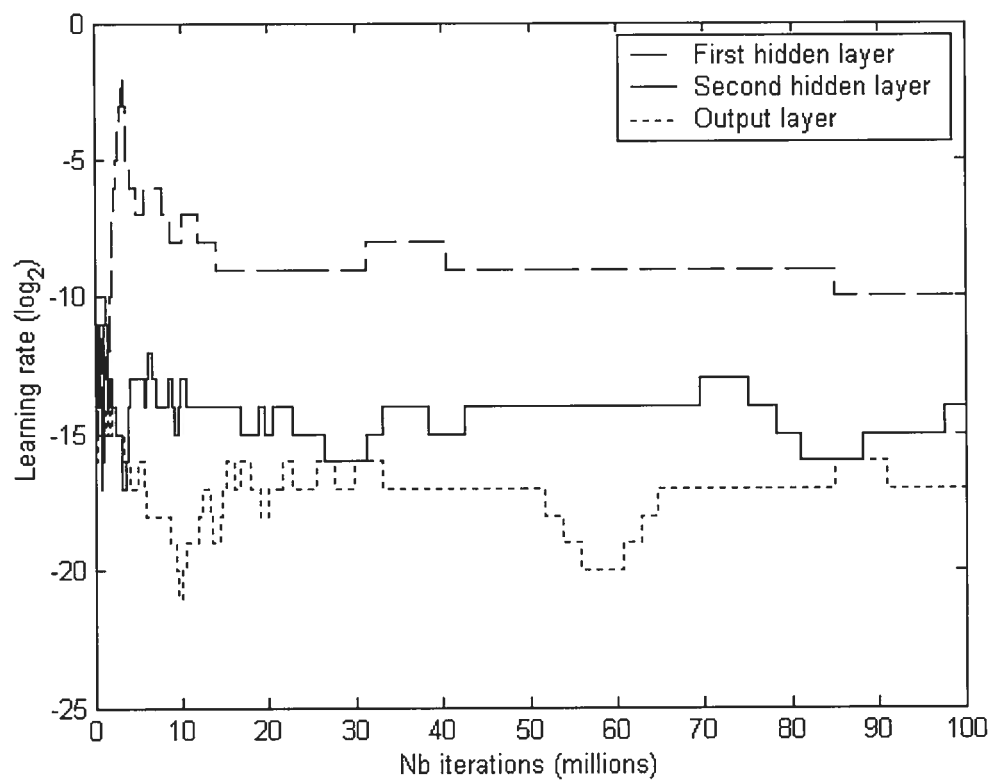


Figure 7. Evolution of the LR for each layer throughout the learning process using the GNDAM.

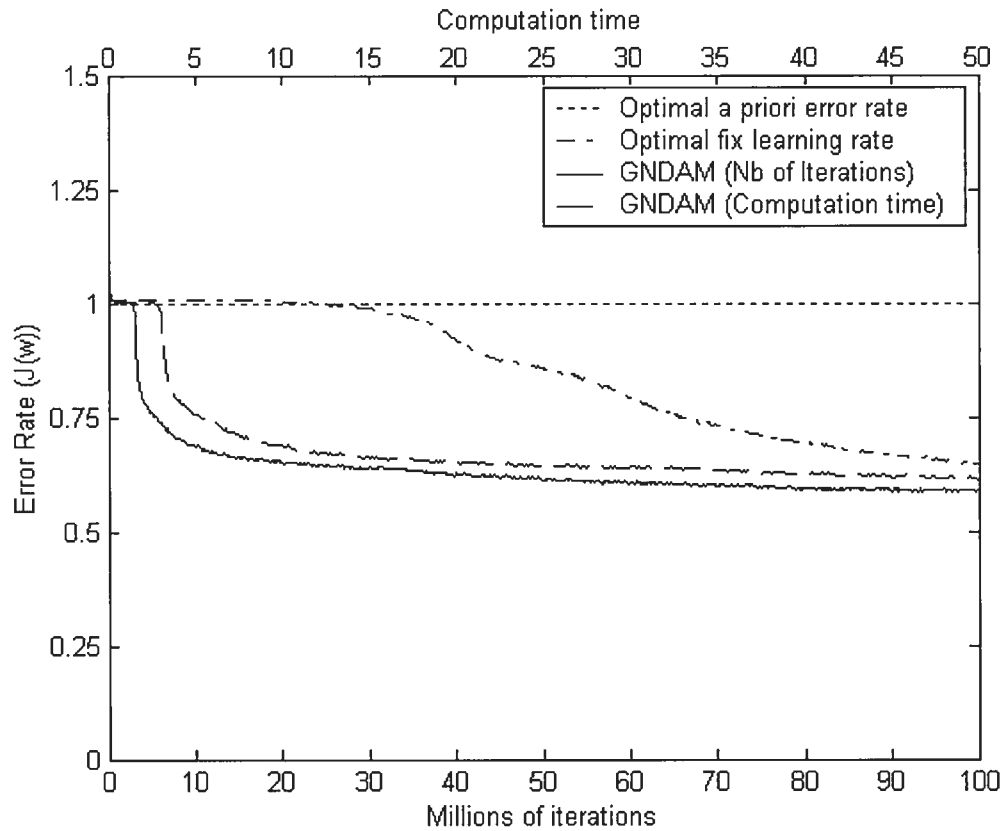


Figure 8. Performance comparison between the optimal fixed LR (2^{-13}) and the GNDAM. The performance of the GNDAM is presented in two different ways. The first one (the curve with the lowest error rate) corresponds to the AM over one hundred million iterations. The second one (computation time), is the GNDAM over fifty million iterations that was stretched over one hundred million iterations. The reason for presenting this second curve is that the GNDAM is computationally two times slower than standard BP since two weight sets are used. Therefore, one iteration of this method equals to two BP iterations in computation time.

GENERAL DISCUSSION

In the introduction it has been mentioned that BP variations may not be worth implementing since at least one of their parameters may be just as hard to tune as the LR using standard BP. In the present study, we introduced a task (the texture task) where standard BP revealed to be impractical using large images because of the LR tuning complexity. Since no better results were obtained using different BP variations, we were forced to develop a new one using a different approach. Instead of using the information about the previous steps of a given weight to adjust its LR, we based our approach on a trial and error heuristic.

The parameters of this new AM should be easy to tune. The only two parameters that should require special tuning are the initial LRs and the prior threshold pdf, and both adjust themselves as learning progresses.

Throughout the present study, network settings (number of units, number of layers, activation function, image size...) were not modified. Modifying one of those parameters would probably have a direct impact on the optimal fixed LR. Using standard BP, this modification would require additional parameter tuning. However, with the GNDAM, this modification should not cause any additional tuning.

As seen above, the GNDAM requires twice the computation time and is more complicated to implement. Therefore, this new AM may not always be the best choice. When the LR is easy to tune, standard BP may be simpler and faster. However, for a large network with slow learning (when many iterations are required before the network learns significantly), the GNDAM may be a good choice. The texture task is a good example of a complicated task: it requires a large network, the data set is infinite and many iterations are required before the error drops under the OAPER. Therefore, an AM that requires many iterations to adapt itself can still be a good alternative.

This paper presents a systematic approach that facilitates LR tuning for complex tasks such as discriminating second-order texture patterns. We believe that the GNDAM for large-slow-learning neural networks that we have developed will prove useful for this complex learning task and that future experiments will assess its robustness to other complex tasks.

REFERENCES

Alpsan, D., Towsey, M., Ozdamar, O., Tsoi, A. & Ghista, D. N. (1995). Are Modified Back-propagation Algorithms Worth the Effort ? *Neural Networks*, **8**, 945-962.

Duda, R.O., Hart, P.E. & Stork, D.G. (2001). Pattern Classification. New York: Wiley. 2nd edition, pp. 282-349.

Jacobs, R. A. (1988). Increased Rates of Convergence Through Learning Rate Adaptation. *Neural Networks*, **1**, 295-307.

Reidmiller, M. & Heinrich, vB (1993). A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. 1, 586-591. New York, IEEE. IEEE International Conference on Neural Networks (San Francisco).

Rumelhart, D.E., Hinton, G.E. & Williams, R.J.. (1986). Learning Internal Representations by Error Propagation. In D.E. Rumelhart, J.L. McClelland & The PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 1, Foundations)*. Cambridge, Massachusetts, MIT Press.

Tollenaere, T. (1990). SuperSAB: Fast Adaptive Back Propagation with Good Scaling Properties. *Neural Networks*, **3**, 561-573.

Watson, A. B. & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

Weibull, W. A. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 292-297.

ANNEX: Pseudo-code of the GNDAM

The present annex has the goal of clarifying the GNDAM. Before presenting the pseudo-code, we begin by defining the constants, functions and variables that will be used.

Constants. Four constants has to be defined: G , K , α , and λ . G is the number of groups in which the weights are separated (each weight group has its own LR). K is a constant that separates two compared LRs, we recommend $K=2$. α and λ are two decayed factors used to exponentially average $P[W(\mathbf{e}_{\text{low}}) < W(\mathbf{e}_{\text{high}})]$ and the pdf of the ideal number of iteration (see experiment 3 for more details). We recommend setting both of them to 0.9.

Functions. $W(\mathbf{x})$ returns the exponentially weighted average of the vector \mathbf{x} . $EW\text{Ftest}(\mathbf{x})$ is the a function representing the *EFW* test as described above. If $W(\mathbf{x})$ is significantly different from 0 this function returns TRUE, otherwise it returns FALSE. $\Psi(x)$ is the psychometric function used by the FQUEST (see experiment 3 for more details). We recommend basing this function on the Weibull distribution. Note that in the function $\Psi^{-1}(y)$, y may not be apart of the domain of Ψ^{-1} and therefore should be truncated to fit this domain.

Initializing variables. The vector \mathbf{w}_{ig} represents the weights of the g^{th} group of the i^{th} weight set ($i = \text{low, high}$ and $g = 1, 2, \dots, G$). Therefore, by initializing the weight set \mathbf{w}_{low} , we do not only initialize the weight values, we also identify each weight to a group. We recommend setting one group by layer. The weight values should be initialized the same way as when using standard BP. η_{ig} corresponds to the LR of its respective weight group (\mathbf{w}_{ig}). Indeed, initializing $\boldsymbol{\eta}_{\text{low}}$ sets the initial LR values. \mathbf{pdf}_g represents the threshold pdf of the g^{th} group (the ideal number of iterations to compare two LRs in log units). p_g represents the online estimation of $P[W(\mathbf{e}_{\text{low}}) < W(\mathbf{e}_{\text{high}})]$ of the g^{th} group. We recommend initializing all elements of vector \mathbf{p} to 1.

Pseudo-code comments:

Line 2 to 4. Simple initializations.

Line 6 to 10. Changes the LR of a given group.

Line 11. FQUEST gives its best estimation of the desired number of iterations (log units).

Line 12 to 16. Apply standard BP to both weight sets and keep track of all the error rates produced (\mathbf{e}).

Line 17 to 30. Keep only the desired weight set and LR.

Line 31 to 35. FQUEST modifies the pdf of the desired number of iteration.

1. **begin initialize** w_{low} , η_{low} , pdf, p
2. $w_{high} \leftarrow w_{low}$
3. $\eta_{high} \leftarrow \eta_{low}$
4. $d_g \leftarrow \text{INCREASE}$, $g = 1, 2, \dots, G$
5. **do for** g from 1 to G **do**
6. **if** $d_g = \text{INCREASE}$ **then**
7. $\eta_{low\ g} \leftarrow \eta_{high\ g} \cdot K$
8. **else**
9. $\eta_{high\ g} \leftarrow \eta_{low\ g} / K$
10. **end**
11. $n \leftarrow$ the threshold of pdf_g
12. **for** t from 1 to $\exp(n)$ **do**
13. select a pattern (set the input layer and the target vector)
14. $e_{it} \leftarrow J(w_i)$, $i = \text{low, high}$
15. apply BP on w_{ih} with η_{ih} , $i = \text{low, high}$ and $h = 1, 2, \dots, G$
16. **end**
17. **if** $W(e_{low}) < W(e_{high})$ and $\text{EWFtest}(e_{low} - e_{high})$ **then**
18. $w_{high} \leftarrow w_{low}$
19. $\eta_{high\ g} \leftarrow \eta_{low\ g}$
20. $d_g \leftarrow \text{DECREASE}$
21. **else**
22. $w_{low} \leftarrow w_{high}$
23. $\eta_{low\ g} \leftarrow \eta_{high\ g}$

```

24.       $d_g \leftarrow \text{INCREASE}$ 
25.      end
26.      if  $W(\mathbf{e}_{\text{low}}) < W(\mathbf{e}_{\text{high}})$  then
27.           $p_g \leftarrow \alpha \cdot p_g$ 
28.      else
29.           $p_g \leftarrow \alpha \cdot p_g + 1 - \alpha$ 
30.      end
31.      if  $\text{EWFtest}(\mathbf{e}_{\text{low}} - \mathbf{e}_{\text{high}})$  then
32.           $pdf_{ii} = pdf_{ii}^{\lambda} \cdot \Psi(n - i + \Psi^{-1}(0.5 / p_g)), \forall i$ 
33.      else
34.           $pdf_{ii} = pdf_{ii}^{\lambda} \cdot (1 - \Psi(n - i + \Psi^{-1}(0.5 / p_g))), \forall i$ 
35.      end
36.      end
37.      until stopping criterion
38.  end

```

Chapitre III

ARTICLE 2

Simulating age-related neurobiological dysfunctions with artificial neural networks: Effect on simple and “Gestalt-like” visual perception

Rémy Allard, Jocelyn Faubert, Ph.D.

ABSTRACT

With aging, complex tasks seem to be more affected than simple tasks. In the second study, neurobiological alterations (NBAs) were simulated with the goal of identifying those that correlated with the perceptive visual deficits. Several NBAs such as synapse loss, myelin sheath degradation and degeneration of intra-cortical inhibition, showed a good correlation with the psychophysical data on aging. This data corresponds to the performance of a group of young adults and a group of elderly for a simple and a complex visual tasks. Habak and Faubert (2000) showed that the thresholds of detections of elderly increase by 29 % and 77 % for stimuli of first and second order respectively, this in comparison with the thresholds of the young adults. Simulations showed that several NBAs hypotheses correlated with these results. However, these various hypotheses do not predict the same performances for other age groups that were not tested by Habak and Faubert (2000). For adults younger than 70.75 years-old, certain hypotheses predicted that the threshold increase ratio of first and second order would be constant and others predicted that the second order would be more affected. For subjects over 70.75, certain hypotheses also predicted a constant ratio and others predicted that the first order would be more affected.

INTRODUCTION

Aging adults often show subtle cognitive and perceptual impairments (Fraik & Jennings, 1992; Faubert, 2002). Knowing what causes these deficits is essential to prevent or reverse their development. Neuroscientists have been trying, in the last decade, to find neurobiological alterations (NBAs) that correlate with these deficits (Peters, Rosene, Moss, Kember, Abraham, Tigges, & Albert, 1996). However, several anatomical and physiological changes occur during the normal course of aging, which makes it possible that some may correlate with perceptual/cognitive impairments without being the underlying cause.

Visual perception and aging

Most of the studies on the functional consequence of normal aging vision have focused either on sensory capacities (i.e. acuity, contrast sensitivity) or higher-level cognitive abilities (Fraik et al., 1992; Kline & Scialfa, 1996). However, there are a number of recent studies that have attempted to elucidate the effect of the normal aging process on perceptual functions (Bennet, Sekuler, McIntosh, & Della-Maggiore, 2001; Faubert & Bellefeuille, 2002; Habak & Faubert, 2000; Herbert, Overbury, Singh, & Faubert, 2002; Sara & Faubert, 2000). Faubert (2002) concluded that what predicts whether normal aging will produce a deficit or change in perceptual function is the complexity of the neural network necessary to achieve the visual task. In other words, all higher-order or 'Gestalt-like'

perceptual tasks such as symmetry perception, 2nd order processing, or integration across visual attributes or spatio-temporal integration will be affected by aging because they all require integration over complex or large neural networks (Faubert, 2002).

The purpose of the present study was twofold. First, we wanted to develop a neural network that was capable of discriminating simple and complex perceptual tasks. The second goal was to simulate potential age-related neurobiological alterations and see whether the neural network output would resemble what was obtained from real psychophysical measurements in the normal young and aging observers. For the present study, we chose to focus on the results of the Habak and Faubert (2000) study on perception of 1st and 2nd order images. There is converging evidence from visual psychophysics modeling, functional imaging, and neuropsychology that 2nd order processing, as defined in the present study, is a more complex perceptual task than 1st order processing (Bertone & Faubert, 2003; Habak et al., 2000; Wilson, 1998). Here we define 1st and 2nd order stimuli as luminance and contrast modulation of texture respectively (see Figure 1 for examples). Habak and Faubert (2000) compared detection thresholds for young (mean age of 22.75 years) and elderly (mean age of 70.75 years) observers. When comparing these two groups, they found that the 1st order and 2nd order detection thresholds increased by 29% and 77% respectively with aging. They argued that the different complexity levels of the stimuli account for the differences in sensitivity losses. They claim that a more complex task should require more

computation and thereby a larger or more complex network, which would explain a greater deficit caused by a subtle widespread neurobiological dysfunction.

Aging visual cortex: NBA hypotheses

Neuroscientists previously held the view that neuronal loss was the major factor causing cognitive impairments in the elderly (Brody, 1955; Devaney & Johnson, 1980; Henderson, Tomlinson, & Gibson, 1980). Recently however, researchers are beginning to adopt a different perspective suggesting that there is no neuronal loss in the visual cortex (for a review see Peters et al., 1996). If this view is correct and the cognitive deficits are not caused by neuronal loss, then some other neurobiological anomaly must cause these deficits. In the recent literature, we found three major neurobiological anomalies that are associated with aging. 1) The myelin sheath breakdown along some axons causing slower signal transmissions and desynchronization of neuron firing (Nielsen & Peters, 2003; Sloane, Hinman, Lubonia, Hollander, & Abraham, 2003). 2) A loss of synapses has been reported in the aged visual cortex (Gandy, 2003; Peters, 2002; Peters & Sethares, 2002; Poe, Linville, & Brunso-Bechtold, 2001). And 3) the major inhibitory neurotransmitter γ -aminobutyric acid (GABA) seems to degrade with aging (Schmolesky, Wang, Pu, & Leventhal, 2000).

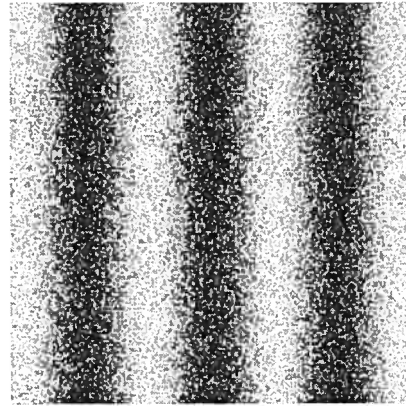
There is a debate on which neurobiological anomaly is responsible for the age-related losses, however the three hypotheses enumerated above are among the

most cited. Although these NBAs probably all occur with aging and that more than one may correlate with sensitivity losses they may not all be responsible for perceptual/cognitive declines. In the present study, we simulated different NBA hypotheses that could cause 1st and 2nd order sensitivity losses with aging.

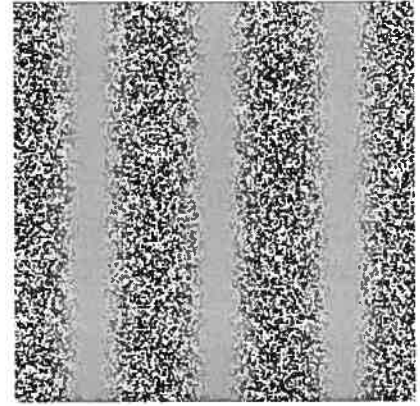
An artificial model

Artificial neural networks (ANNs) are often used to model the brain since they are functionally analogous: simple functional elements (neurons/units), once properly linked together accomplish complex tasks. In the present study, an ANN simulated age-related sensitivity losses to 1st and 2nd order stimuli by NBAs. Based on the hypothesis that ANN and the brain are functionally analogous, if the NBA mostly responsible for visual sensitivity deficits is simulated in an ANN, the ANN losses should be similar to the one caused by normal aging (29% and 77% detection threshold increase for 1st and 2nd order stimuli respectively).

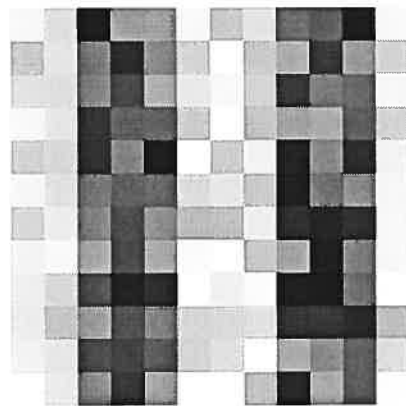
Figure 1a-b shows examples of the stimuli used by Habak and Faubert (2000). The ANN had to learn to detect the orientation (horizontal or vertical) of similar stimuli (figure 1c-d, see methodology below for more details). The detection threshold was the modulation (luminance or texture) necessary to detect the orientation.



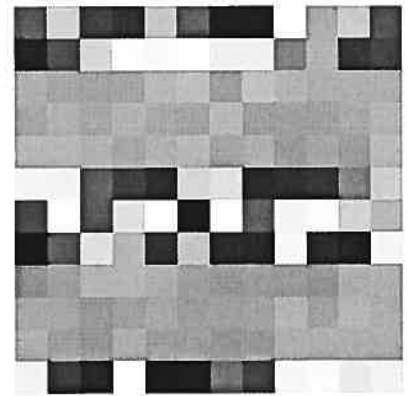
(a)



(b)



(c)



(d)

Figure 1. (a) and (b) are stimuli similar to the one used by Habak and Faubert (2000). (a) is a 1st order stimulus (a grating defined by luminance with added noise) and (b) is a 2nd order stimuli (a grating defined by contrast). (c) and (d) are similar to (a) and (b) respectively except that a square wave is used instead of sinus and the image size is reduced since less pixels are used to define the images.

METHODOLOGY

The double task approach

The research methodology utilized in this study can be summarized as follows: First we created an ANN that was able to discriminate the orientation of 1st and 2nd order stimuli. After the learning process, the ANN performance should be optimal (or close) and thereby simulate the performance of the youngest group. Then, for each NBA hypothesis (e.g.: synapse loss), aging was simulated according to both tasks. To do so, the required distortion quantities (α_1 for 1st and α_2 for 2nd order) introduced in the network causing the 1st and 2nd order detection thresholds to increase in the same proportion as normal aging (29% and 77% respectively) was found. If $\alpha_1 \approx \alpha_2$ the simulated NBA hypothesis is said to strongly correlate with aging psychophysics data for those two tasks. Otherwise there is only a weak correlation and the hypothesis is not likely to cause, at least by itself, the aging losses for processing 1st and 2nd order stimuli.

Stimuli

The first methodological step was to create an ANN able to detect the orientation of 1st and 2nd order defined patterns. If the stimuli all had the same spatial frequency and phase, the ANN could have easily utilized a simple heuristic to classify the patterns (the same pixel zone would always correspond to the same

intensity (luminance or texture) for a given orientation). Therefore, the stimuli must be conceived to force the ANN to generalize its pattern recognition to simulate true orientation perception. Consequently, patterns had different periods, phases and amplitudes.

Because of the computation time constraint, the image size had to be reduced. Therefore, square wave patterns were used instead of sin wave patterns (figure 1b). Two possible periods, where one was not a factor of the other, characterized the patterns. If one period had been the factor of the other, two pixels at a distance of the largest period must always be of the same intensity. To satisfy these conditions, 12-by-12 pixel images representing square wave patterns with periods of either 4 or 6 pixels were created. The 1st order stimuli were created using the following function:

$$L(x, y) = 0.5 \cdot \text{rand}(x, y) + 0.5 \cdot \begin{cases} I_1 & \text{if } \text{mod}(z, p) < p/2 \\ I_2 & \text{otherwise} \end{cases}$$

$L(x, y)$ represents the normalized luminance value of the pixel located at the x-y position. $\text{rand}(x, y)$ returns a random value between 0 and 1 for each pixel. $\text{mod}(a, b)$ returns the remainder of b divided by a. I_1 and I_2 are the two square wave intensities randomly chosen between 0 and 1 for each pattern. p is the pattern period (4 or 6) and, z depended on the pattern orientation (=y for horizontal and =x for vertical).

The 2nd order patterns were created by the following function:

$$L(x, y) = 0.5 + (\text{rand}(x, y) - 0.5) \cdot \begin{cases} I_1 & \text{if } \text{mod}(z, p) < p/2 \\ I_2 & \text{otherwise} \end{cases}$$

Creating the stimuli as mentioned above demonstrates the complexity of the stimuli used. Even the 1st order stimuli had a certain complexity level. Detecting a pattern orientation in noise when the period, phase and amplitude are all randomly selected is complex for an ANN. Adding to this that the ANN did not know which of the patterns was shown (1st or 2nd order) demonstrates the complexity of the task the ANN had to learn.

Network architecture

In our simulation, a unit (artificial neuron) represented a group of similarly configured neurons (O'Reilly & Munkata, 2000). Therefore, the activation function of a unit was a sigmoid where the minimum and maximum values were 0 and 1. The weights representing excitatory connections were truncated between 0 and 1, and the inhibitory ones were truncated between -1 and 0. Since most brain neurons all have post-synaptic connections of the same type (excitatory or inhibitory) (O'Reilly et al., 2000), the post-synaptic weights of a given unit were of the same type. The inhibitory units modeled lateral inhibition.

Figure 2 shows the network architecture. Obviously, there were 144 input units (a 12-by-12 pixel image). The first excitatory hidden layer (288 units) received projections from the input layer directly and via the first inhibitory hidden layer

(144 units). The second excitatory hidden layer (144 units) received projections from the first excitatory hidden layer directly and also via the second inhibitory hidden layer (72 units). The second excitatory hidden layer projected directly to the output layer. The target values of the two output units were either [0 1] or [1 0] depending on the stimulus orientation (input layer). A bias unit was also connected to each non-input unit. Learning was performed using the Generalized no-decrease adaptive method (Allard & Faubert, In press) and was stopped after two hundred million iterations.

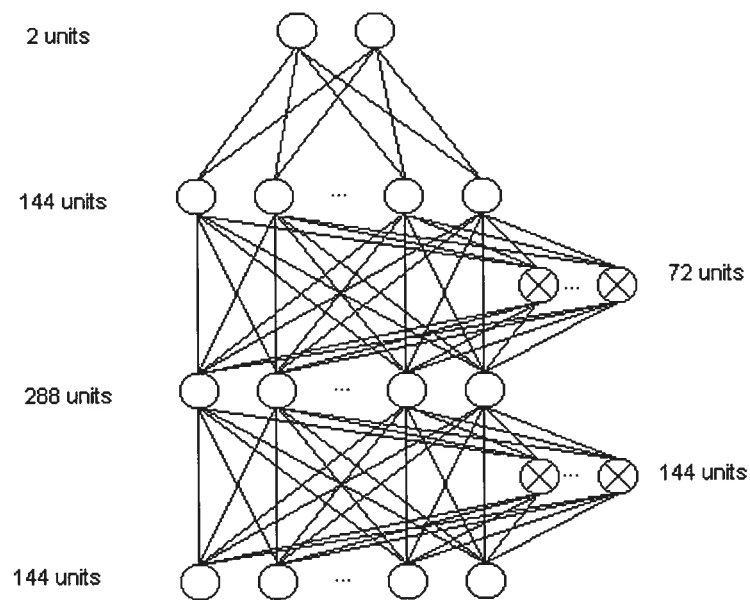


Figure 2. ANN architecture. The inhibitory units are marked by a 'X'. A bias was also connected to each non-input unit (do not appear on the figure).

Simulation hypothesis

Each of the simulation hypotheses mentioned bellow was applied to excitatory, inhibitory or both types of units or connections. The distortion quantities that simulated 1st and 2nd order sensitivity losses were found for each simulation hypothesis. Since little is known about NBAs that occur with aging, the simulated hypotheses were not limited to the ones found in the literature. Therefore, to cast a wider range of possible biological alterations, a variety of artificial alterations were performed even if some did not have present anatomically or physiologically support in the literature.

Synapse loss

Recent studies have reported a loss of synapse in the visual cortex attributed to aging (Gandy, 2003; Peters, 2002; Peters et al., 2002; Poe et al., 2001). Removing connections between units simulated this biological hypothesis. For this simulation, the distortion quantity introduced in the network was the number of connections removed. Removing a connection consisted of setting its weight value to 0. Since it is generally accepted that the less useful synapses are the first ones to be lost (Poe et al., 2001). The weaker connections were the first to be removed.

Neurotransmitter alteration

Weight connections between units modeled the neurotransmitters. In our network there were two types of connections (excitatory and inhibitory) simulating the two

major neurotransmitters glutamate (excitatory) and GABA (inhibitory). The weight connection alteration hypothesis was divided in three possible alterations. The weight connections could either be all decreased, all increased, or some increased and others decreased (noise condition). The increase and decrease conditions simulated a neurotransmitter gain or lost respectively. In the literature, a decrease of the GABA neurotransmitter has been reported in the visual cortex (Schmolesky et al., 2000) and glutamate level seems to remain constant with aging (Peters, 2002). The noise condition simulated a non-uniform alteration in the communication between units. A biological equivalence for this condition could be a non-uniform alteration at the neurotransmitter level, which could result in synaptic noise.

The weights were modified using a soft weight bounding technique, which prevented them to be smaller than 0 or greater than 1 (O'Reilly et al., 2000):

$$w_i = w_i + [x_i]_+ * (1 - w_i) + [x_i]_- * w_i$$

where $x_i = \alpha$ for the increase condition, $x_i = -\alpha$ for the decrease condition, $x_i = \text{gauss}(\alpha)$ for the noise condition (gauss returns a random value following a Gaussian distribution of mean 0 and standard deviation α) and α represents the distortion quantity introduced in the network ($\alpha > 0$). The operator $[x]_+$ returns x if $x > 0$ and 0 otherwise and $[x]_-$ returns x if $x < 0$ and 0 otherwise.

Firing signal alteration

In this hypothesis, the output of hidden units (o_i) (either excitatory units, inhibitory units or both) were modified by the following function:

$$o_i = o_i * (1 + x_i)$$

x_i being one of the three conditions enumerated in the neurotransmitter hypothesis.

Since a unit simulates a population of similarly configured neurons, introducing noise at the unit firing could be considered as simulating desynchronized neuron firing. Therefore, the myelin sheath breakdown was simulated indirectly since we assumed its damage would result in a noisy signal from a population of similarly configured neurons (units).

Although there seems to be a consensus that there is no significant neuronal loss with aging, a recent study advanced the hypothesis that inhibitory neurons may be lost with aging (Hof, Nimchinsky, Young, & Morrison, 2000). The authors argued that since there is less inhibitory neurons than excitatory ones, a significant inhibitory neuronal loss may not cause a global significant neuronal loss. This last hypothesis was simulated by the activation-decrease condition, which simulated a uniform neuronal loss.

Threshold evaluation

The threshold criterion was 75% of correct response. To evaluate a detection threshold, a constant stimulus method was used where each intensity (contrast amplitude) was computed one thousand times. Afterwards, bootstrapping was

computed to fit a Weibull distribution function on these results to approximate threshold (Weibull, 1951).

RESULTS

Learning

Figure 3 shows the ANN learning curve (error rate) for 1st and 2nd order stimuli perception. This figure reveals that for the ANN the 2nd order patterns were, by far, harder to learn than the 1st order ones. Indeed, it required more iterations before the error rate dropped below the optimal a priori error rate ($=0.25$). We can also notice that the two curves reached an asymptote suggesting that the network found a minimum.

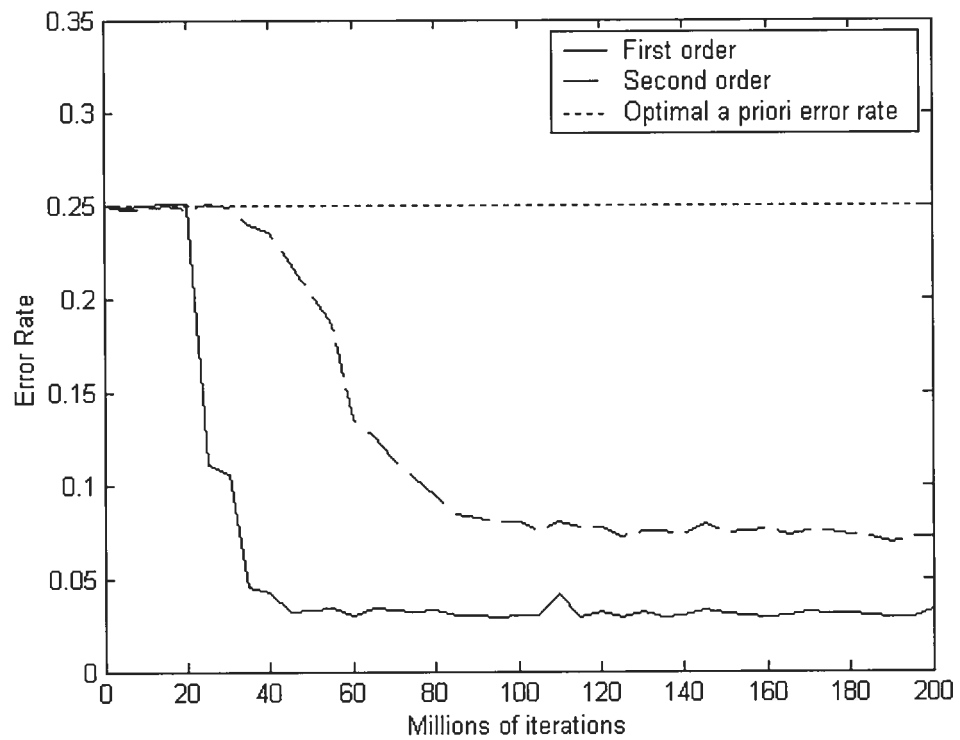


Figure 3. Learning curves of the artificial neural network. The optimal a priori error rate is the minimized error rate when nothing is known about the input stimulus (chance rate).

Hypothesis results

Table 1 summarizes the results obtain when distortion was introduced in the network. The value presented for each condition was calculated using the following function:

$$\text{ratio} = \log_2 (\alpha_1 / \alpha_2)$$

where α_1 and α_2 represent the distortion quantity (condition dependent) introduced in the network that caused a 29 % 1st order detection threshold increase (α_1) and a 77 % 2nd order detection threshold increase (α_2). A ratio near 0 means that, for the given condition, the same distortion quantity made the detection threshold increase in the same proportion as normal aging (1st order: 29 %; 2nd order: 77 %). A ratio<0 implies that the 1st order sensitivity was more affected compared with normal aging deficit and a ratio>0 implies that the condition affected more the 2nd order sensitivity.

Hypothesis	Excitatory	Inhibitory	Both
Connection-removal	0.01	-0.76	0.15
Connection-noise	-0.86	1.06	-0.34
Connection-increase	-0.52	1.00	0.27
Connection-decrease	-2.07	0.05	-0.42
Unit activation-noise	0.06	0.12	0.13
Unit activation-increase	-0.08	-1.70	0.84
Unit activation-decrease	-1.93	0.13	-0.42

Table I. Results of each NBA simulated. A value near 0 represents that the simulated alteration correlates well with the psychophysics data (a given distortion quantity inserted in the network made the detection threshold increase of 29% and 77% for the 1st order and 2nd order stimuli respectively).

The connection loss, which simulated the decrease in the number of synapses in the aging visual cortex, correlated if only excitatory (0.01) or both types of synapses are lost (0.15), but not if only inhibitory synapses are lost (-0.76). Those results suggest that the most important inhibitory synapse losses that occur in the brain is not the main cause of the 1st and 2nd order sensitivity losses that occur with aging.

In the connection-noise condition, no alterations (inhibitory, excitatory or both) correlated with the sensibility losses. Therefore, noise alteration at the synapse level did not seem to correlate with 1st order and 2nd order sensitivity losses.

Still in regards to the connection simulation, a decrease of all the inhibitory synapse strength (connection-decrease condition) correlated (0.05). This condition simulated a loss of the GABA neurotransmitter reported in the visual cortex.

The simulations performed on the units activation modified the output signal of a population of similarly configured neurons. As opposed to adding noise at the connection level where no correlation was found, adding noise at the units' output activation level produced a correlation between our simulation and the psychophysics data. This was the case when only the excitatory units were affected (0.06), or only the inhibitory ones (0.12) or both of them (0.13). As mention above, it could be argued that this noise on the output firing average of a

population of neuron could be caused by the myelin sheath breakdown that desynchronizes the neuron firing.

A firing rate increase has been reported in the visual cortex (Schmolesky et al., 2000) and was simulated in our model. Although the authors argued that this increased firing rate was caused by a decrease in the GABA neurotransmitter level (which also correlated in our model), directly increasing the excitatory or decreasing the inhibitory neurons firing rate also correlated (-0.08 and 0.13 respectively).

DISCUSSION

A total of 21 NBAs were simulated. Many of them (8) correlated (≤ 0.15) with the aging sensibility loss of 1st and 2nd order stimuli perception. This means that for 8 simulated conditions out of 21, when a certain distortion quantity caused a 1st order detection threshold increase of 29%, the same network had a 2nd order detection threshold increase of $\approx 77\%$. Many hypotheses correlating with aging perception decline could be seen as supporting the ANN simulation of age-related deficits since they have similar behavior.

Figure 4 shows the detection threshold of 1st and 2nd order conditions as a function of the distortion quantity introduced in the network for the conditions that

correlated with the aging deficits. The 8 graphs were classified in three different groups according to their curve shape. The first group (connection lost for excitatory and, excitatory and inhibitory conditions) represents the curves where the 1st and 2nd order sensitivity loss seems to fit perfectly. These curves suggest that the 1st and 2nd order perceptions are affected in the same proportion one relative to the other throughout aging (while distortion is introduced). In support of the synapse loss hypothesis, our simulation suggests that if the 1st order detection threshold increases by a factor of x , then the 2nd order detection threshold should increase by a factor of $77 \cdot x / 29$. Habak and Faubert (2000) already demonstrated that this equation is true between 22.75 and 70.75 year-old subject groups. Unfortunately, no psychophysics data at this time determines what are the correct proportion changes as a function of other age groups.

The second group (all activation output-noise conditions), has similar curve types as the first group with the exception that when too much distortion is introduced, the $77/29$ proportion becomes invalid and the 1st order condition becomes more affected. One way to evaluate which of the two first groups correlates best with the psychophysics data would be to test aged people over the entire life span and see which curve reveals itself to be inexact.

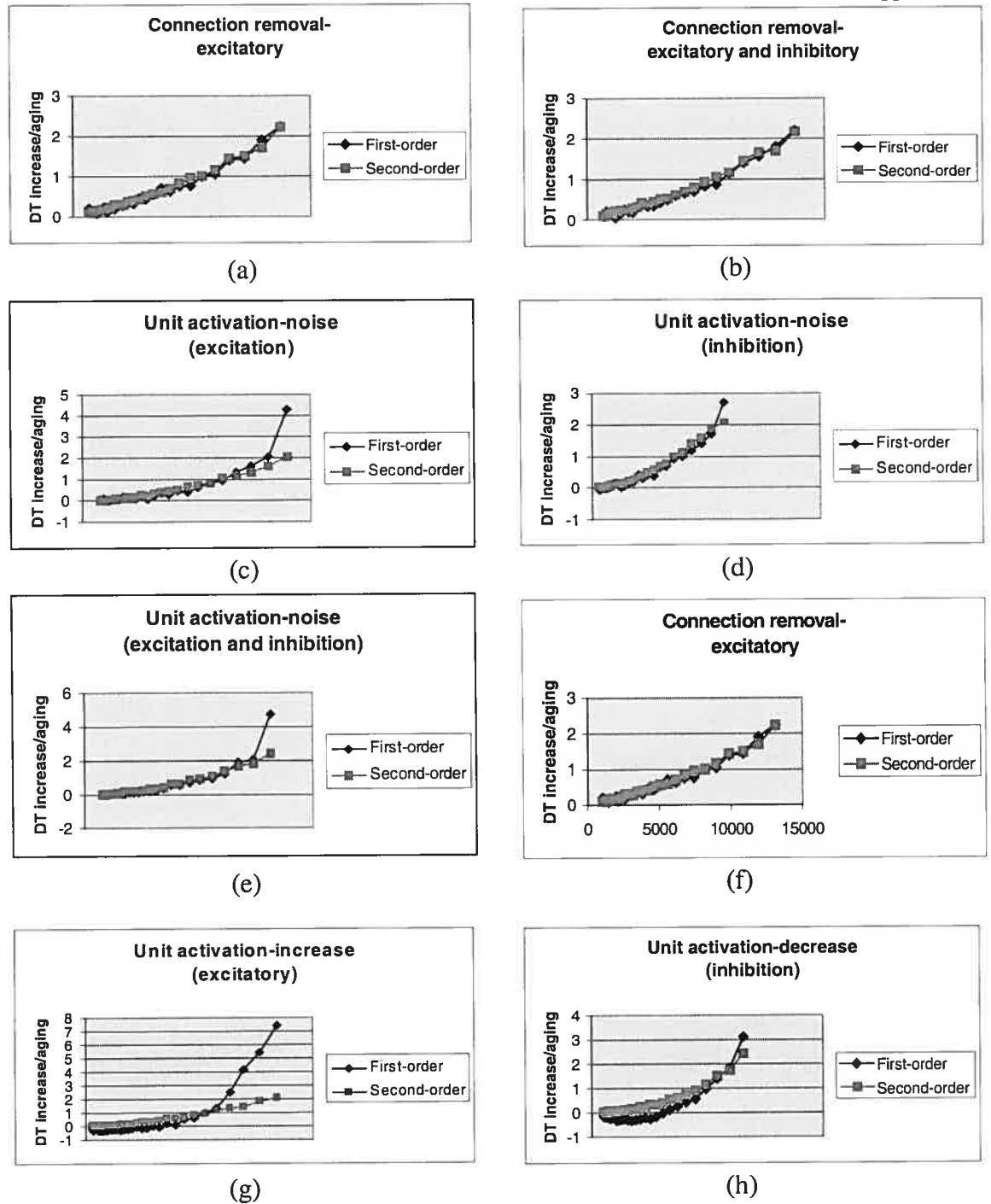


Figure 4. Presents the curve of detection threshold increase in function of the distortion quantity inserted in the network. The detection threshold increase is presented as an aging ratio; the 1st order detection threshold increase percentage was divided by 29% and the 2nd order by 77%.

The last group (inhibitory connection decrease, excitatory activation increase and inhibitory activation decrease) has a very different curve type. For those conditions, when a large distortion quantity is introduced it behaves like the second group (the 1st order sensitivity becomes more affected). However, when a low level of distortion was introduced, a 2nd order sensitivity loss was noted and a sensitivity gain occurred for the 1st order stimuli. How was that possible? Computationally speaking, two solutions are possible. First, the network may have not learned the optimal solution for each task. Obviously, if it had, sensitivity increase would have been impossible. The second possibility is that once the solution was found, a small distortion quantity helps the 1st order sensitivity but would disadvantage the 2nd order one. Therefore, because of its limitation, the network may not have been able to learn the optimal sensitivity for both tasks and would have found a compromise between them. Introducing a small amount of distortion would have disrupted this compromise in favor of 1st order sensitivity.

Each condition that correlated with the psychophysics data was classified in one of the three groups enumerated above based on their 1st order versus 2nd order sensitivity behavior according to the distortion quantity introduced in the network. Interestingly, it is easy to link each group to a NBA hypothesis already mentioned in the introduction. The first group represented synapse loss hypotheses, either only for excitatory or both type of synapses. The second group simulated the noise insertion at the output signal of a similarly configured neuron population that could be caused by a myelin sheath breakdown. The third group may seem

composed of disparate hypotheses. However, all the conditions in this last group and none in the other groups had the effect of increasing the units' excitatory level compared to the inhibitory one. Decreasing the inhibitory connection weights, increasing the unit activation level or decreasing the unit activation level all have in common the effect of reducing the inhibition in the network. Therefore, the 1st order hypersensitivity was caused by diminishing the inhibition/excitation ratio in the network. Interestingly, it has been reported that a hypersensitivity could be caused by certain brain disorders. Autism, for instance, is known to improve the sensitivity of simple tasks but has the opposite effect on more complex tasks (Bertone et al., 2003). This was well simulated in our latest conditions. However, it is not very plausible that this type of distortion is what happens with aging since no one, as far as we know, ever reported a hypersensitivity of certain tasks for the middle-aged or older people.

The simulations performed in the present study had the goal of supporting (or not supporting) certain NBA hypotheses that could occur in the aging visual cortex. Surprisingly, the simulations supported many major hypotheses (synapse loss, myelin sheath breakdown and GABA decrease). However, each one of the hypotheses groups had a different 2nd order / 1st order sensitivity loss ratio curve according to the distortion quantity. Therefore, with more psychophysics data on the effect of age throughout the life span, our simulation would probably clearly support one NBA more likely to be responsible for the age-related sensitivity loss seen in 1st order and 2nd order perception.

REFERENCES

Bennet, P. J., Sekuler, A. B., McIntosh, A. R. & Della-Maggiore, V. (2001). The effects of aging on visual memory : Evidence for functional reorganisation of cortical networks. *Acta Psychologica*, **107**, 249-273.

Bertone,A. & Faubert,J. How is complex second-order motion processed? Vision Research . 2003.

Ref Type: In Press

Brody, H. (1955). Organization of the cerebral cortex. III. A study of aging in the human cerebral cortex. *Journal of Computational Neurology*, **102**, 511-556.

Craik,F.I.M. & Jennings,J.M. 1992. Human memory. In: *Handbook of Aging and Cognition* (Ed. by F.I.M.Craik & T.A.Salthouse), pp. 51-110. Hillsdale, NJ: Lawrence Erlbaum.

Devaney, K. O. & Johnson, H. A. (1980). Neuron loss in the aging visual cortex of man. *J Gerontol A Biol Sci Med Sci*, **35**, 836-841.

Faubert, J. (2002). Visual Perception and Aging. *Canadian Journal of Experimental Psychology*, **56**, 164-176.

Faubert, J. & Bellefeuille, A. (2002). Effects of aging on intra- and inter-attribute spatial frequency information for luminance, color, and working memory. *Vision Research*, **42**, 369-378.

Gandy, S. (2003). Estrogen and Neurodegeneration. *Neurochemical Research*, **28**, 1003-1008.

Habak, C. & Faubert, J. (2000). Larger effect of aging on the perception of high-order stimuli. *Vision Research*, **40**, 943-950.

Henderson, G. B., Tomlinson, E. & Gibson, P. H. (1980). Cell counts in human cerebral cortex in normal adults throughout life using an image analyzing computer. *J Neurol Sci*, **46**, 113-136.

Herbert, A. M., Overbury, O., Singh, J. & Faubert, J. (2002). Aging and bilateral symmetry detection. *Journal of Gerontology B: Psychological Sciences*, **57**, P241-P245.

Hof, P. R., Nimchinsky, E. A., Young, W. G. & Morrison, J. H. (2000). Numbers of Meynert and Layer IVB Cells in Area V1: A Stereologic Analysis in Young and Aged Macaque Monkeys. *The Journal of Comparative Neurology*, **420**, 113-126.

Kline, D. W. & Scialfa, C. T. 1996. Visual and auditory aging. In: *Handbook of the Psychology of Aging* (Ed. by J. E. Birren & K. W. Schaie), pp. 181-203. Academic Press Inc.

Nielsen, K. & Peters, A. (2003). The effects of aging on the frequency of nerve fibers in rhesus monkey striate cortex. *Neurobiology of Aging*, **21**, 621-628.

O'Reilly, R.C. & Munkata, Y. 2000. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press edn.

Peters, A. (2002). Structural changes that occur during normal aging of primate cerebral hemispheres. *Neuroscience and Biobehavioral Reviews*, **26**, 733-741.

Peters, A., Rosene, D. L., Moss, M. B., Kember, T. L., Abraham, C. R., Tigges, J. & Albert, M. S. (1996). Neurobiological bases of age-related cognitive decline in the rhesus monkey. *Journal of Neuropathology and Experimental Neurology*, **55**, 861-874.

Peters, A. & Sethares, C. (2002). The Effects of Age on the Cells in Layer 1 of Primate Cerebral Cortex. *Cerebral Cortex*, **12**, 27-36.

Poe, B. H., Linville, C. & Brunso-Bechtold, J. (2001). Age-Related Decline of Presumptive Inhibitory Synapses in the Sensorimotor Cortex as Revealed by the Physical Disector. *The Journal of Comparative Neurology*, **439**, 65-72.

Sara, M. & Faubert, J. (2000). Aging, perception and visual short-term memory for luminance defined form. *Ophtalmic and Physiological Optics*, **20**, 314-322.

Schmolesky, M. T., Wang, Y., Pu, M. & Leventhal, A. G. (2000). Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nature Neuroscience*, **3**, 384-390.

Sloane, J. A., Hinman, J. D., Lubonia, M., Hollander, W. & Abraham, C. R. (2003). Age-dependent myelin degeneration and proteolysis of oligodendrocyte proteins is associated with the activation of calpain-1 in the rhesus monkey. *Journal of Neurochemistry*, **84**, 157-168.

Weibull, W. A. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, **18**, 292-297.

Wilson, H.R. 1998. In: *Cerebral Cortex: Models of Cortical Circuitry* (Ed. by P.S. Ulinski & E.G. Jones), pp. 14. New York, Plenum.

Chapitre IV

DISCUSSION GÉNÉRALE DES RÉSULTATS

Dans la première étude, un algorithme d'apprentissage pour les réseaux de neurones artificiels a été développé. Ce nouvel algorithme basé sur la rétropropagation (*backpropagation*) a comme but d'ajuster le taux d'apprentissage. Beaucoup d'autres algorithmes ayant le même but existent déjà. Comme vu dans l'introduction, il est possible de diviser ceux-ci en deux groupes distincts. Le premier utilise une approche globale, c'est-à-dire que tous les poids ont le même taux d'apprentissage et que ce dernier est ajusté en fonction de l'erreur produite par le réseau. Le second groupe utilise plutôt une approche locale, c'est-à-dire que chaque poids a son propre taux d'apprentissage qui est ajusté en fonction des dernières modifications apportées à ce poids.

Le *Generalized no-decrease adaptive method* (GNDAM) développé dans le premier article se distingue de ces deux approches. Premièrement, il se distingue largement de l'approche locale puisque le taux d'apprentissage n'est pas ajusté en fonction de la direction des changements des poids et chaque poids n'a pas son propre taux d'apprentissage. Il se différencie également de l'approche globale puisque tous les poids n'ont pas le même taux d'apprentissage. En effet, les poids sont plutôt divisés en groupe, par exemple un taux d'apprentissage par couche

d'unités, et chaque groupe a son propre taux d'apprentissage. Par contre, le GNDAM, comme l'approche globale, se base sur l'erreur pour ajuster le taux d'apprentissage. Pour ce faire, il utilise une nouvelle approche basée sur une heuristique d'essais et erreurs qui compare la performance de deux taux d'apprentissage tandis que l'approche globale ajuste le taux d'apprentissage en se basant sur l'erreur produite par un seul réseau de neurones artificiels. Le GNDAM est donc fondamentalement différent de toutes les autres méthodes adaptatives ajustant le taux d'apprentissage.

Puisque le GNDAM utilise une nouvelle approche, de futures recherches s'avèrent nécessaires pour optimiser son efficacité. Premièrement, il est essentiel de le comparer avec d'autres algorithmes sur une variété de tâches de façon à déterminer ses forces et ses faiblesses. De plus, il y a des possibilités d'améliorer le GNDAM. Par exemple, une des limites du GNDAM est que les poids n'ont pas leur propre taux d'apprentissage. Une façon de remédier à cet inconvénient serait de fusionner le GNDAM avec un algorithme de l'approche locale. La plupart des approches locales ont plusieurs paramètres à ajuster dont un plus critique qui influence directement le taux d'apprentissage. Au lieu d'ajuster directement le taux d'apprentissage, le GNDAM pourrait alors ajuster ce paramètre. Ceci pourrait améliorer l'algorithme de l'approche locale en rendant un de ses paramètres flexible à l'aide du GNDAM. Ce nouveau type d'approche hybride pourrait donc améliorer davantage la vitesse d'apprentissage pour les réseaux de neurones

artificiels utilisant la rétropropagation en rendant un paramètre quelconque flexible.

Le développement du GNDAM a permis les simulations effectuées dans la deuxième étude. Dans cet article, un réseau de neurones artificiels a appris à distinguer l'orientation de stimuli de premier et de deuxième ordres. Ensuite, différentes simulations d'altérations neurobiologiques (ANBs) ont servi à simuler le vieillissement. Leurs effets sur les seuils de détections des stimuli de premier ordre et de deuxième ordre ont été comparés aux résultats de l'étude de Habak et de Faubert (2000). Ceux-ci ont démontré, avec le vieillissement chez l'humain, une augmentation des seuils de détection de 29 % et 77 % pour des stimuli de premier et de deuxième ordres respectivement.

Les résultats de l'étude actuelle ont démontré que plusieurs hypothèses d'ANBs retrouvées dans la littérature corrélaient avec les simulations effectuées. Cependant, ces différentes hypothèses ne prédisent pas les mêmes performances pour des quantités de distorsions différentes, c'est-à-dire qu'ils ne prédisent pas les mêmes déficits au niveau du premier ordre et du deuxième ordre pour d'autres groupes d'âge que ceux testés par Habak et Faubert (2000). Par exemple, si une quantité de distorsion est introduite pour causer une augmentation du seuil de détection des stimuli de deuxième ordre de 39 % ($77 \% / 2$), de combien augmentera le seuil des stimuli de premier ordre ? Certaines hypothèses telles la perte de synapses ou la dégradation de la gaine de myéline prédisent, selon nos

simulations, que ce seuil augmentera d'environ 15 % ($29 \% / 2$), alors que l'hypothèse de la diminution d'inhibition intra-corticale prédit une augmentation beaucoup moindre et peut-être même une diminution, c'est-à-dire une hypersensibilité. En introduisant une quantité de distorsion pour que le seuil de détection du deuxième ordre augmente de 144 % ($77 \% \times 2$), l'hypothèse de pertes de synapses prédit une augmentation du seuil de détection des stimuli de premier ordre d'environ 58 % ($29 \% \times 2$) alors que les hypothèses de la dégradation de la gaine de myéline et de la dégénérescence de l'inhibition intra-corticale prédisent une augmentation supérieure à 58 %. Il serait donc intéressant d'avoir des données sur l'augmentation du seuil de détection des stimuli de premier et de deuxième ordres pour d'autres groupes d'âge. Ces données permettraient de voir quel comportement les seuils ont en fonction du vieillissement. De plus, elles permettraient aux simulations effectuées dans ce présent mémoire d'appuyer certaines hypothèses et d'en rejeter d'autres.

D'autres tâches, affectées par le vieillissement, telles la perception de symétrie et la perception du mouvement pourraient également être simulées. Il serait intéressant de voir si les résultats observés dans le présent mémoire sont reproductibles avec d'autres tâches. Dans un tel cas, l'hypothèse proposant que les réseaux de neurones artificiels soient de bons modèles du cortex et celles sur les ANBs qui corréleraient avec les données psychophysiques seraient renforcées.

De plus, la simulation de d'autres tâches perceptives pourraient permettre de simuler directement la dégradation de la gaine de myéline. Puisque la dégradation de la gaine de myéline influence la vitesse de transmission le long des axones, pour simuler sa dégradation, une dimension temporelle devrait être ajoutée au réseau de neurones artificiels. Une tâche impliquant la perception de mouvement, par exemple, serait appropriée. Un réseau de neurones artificiels récurrent pour traiter le mouvement permettrait donc de simuler une désynchronisation de l'activation des neurones.

Chapitre V

CONCLUSION

Dans la première étude, un nouvel algorithme ajustant le taux d'apprentissage dans les RNAs a été développé. Il a été démontré, en comparaison avec le backpropagation standard, que cet algorithme accélère la vitesse d'apprentissage pour la tâche de discrimination d'orientation de stimuli définis par la texture.

Dans la seconde étude, un RNA a appris à discriminer l'orientation de stimuli définis par la luminance et par la texture simulant la performance de jeunes adultes. Ensuite, différentes ANBs ont été appliquées sur le RNA dans le but de simuler les déficits observés lors du vieillissement normal chez l'humain. Plusieurs ANBs ont corrélé avec ces déficits. Cependant, les différentes ANBs simulées n'ont pas prédit les mêmes déficits pour certains groupes d'âges pour lesquels aucune donnée n'est présentement disponible dans la littérature.

Ce mémoire ouvre la porte à quatre sujets de recherche. Premièrement, déterminer les forces et les faiblesses du GNDAM. Deuxièmement, développer de nouveaux algorithmes d'apprentissages en combinant le GNDAM avec un algorithme d'approche locale. Troisièmement, étudier la progression du vieillissement de tâches simples et complexes au lieu de simplement comparer les pertes entre les

jeunes et les personnes âgées ce qui donnerait de précieux indices sur les causes neurales du vieillissement. Enfin, simuler artificiellement davantage de tâches à l'aide de différents types de réseaux de neurones artificiels afin de valider ou d'invalidier les résultats obtenus dans ce présent mémoire.

BIBLIOGRAPHIE

Bennet, P. J., Sekuler, A. B., McIntosh, A. R. & Della-Maggiore, V. (2001). The effects of aging on visual memory : Evidence for functional reorganisation of cortical networks. *Acta Psychologica*, **107**, 249-273.

Brody, H. (1970). Structural changes in the aging nervous system. *Interdiscipl Top Gerontol*, **7**, 9-21.

Cabeza, R. (1997). Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. *Journal of Neuroscience*, **17**, 391-400.

Cavanagh, P. & Mather, G. (1989). Motion: The long and short of it. *Spatial Vision*, **4**, 103-129.

Cragg, B. G. (1975). The density of synapses and neurons in normal, mentally defective and aging human brains. *Brain* 981-990.

Engel, A. K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, **5**, 16-25.

Faubert, J. (2002). Visual Perception and Aging. *Canadian Journal of Experimental Psychology*, **56**, 164-176.

Gibson, P. H. (1983). EM study of the number of cortical synapses in the brains of ageing people and people with Alzheimer-type dementias. *Neuropathol*, **62**, 127-133.

Hutterlocher, P. R. (1979). Synaptic density in human frontal cortex: developmental changes and effects of aging. *Brain research*, **163**, 195-205.

Jacobs, R. A. (1988). Increased Rates of Convergence Through Learning Rate Adaptation. *Neural Networks*, **1**, 295-307.

Leuba, G. & Kraftsik, R. (1994). Changes in volume, surface estimate, three-dimensional shape and total numbers of neurons of the human primary visual cortex from midgestation until old age. *Anat Embryol*, **190**, 351-366.

Leventhal, A. G., Wang, Y., Pu, M., Zhou, Y. & Ma, Y. (2003). GABA and Its Agonists Improved Visual Cortical Function in Senescent Monkeys. *Science*, **300**, 812-815.

Madden, D. J., Turkington, T. G., Provenzale, J. M., Denny, L. L., Hawk, T. C., Gottlob, L. R. & Coleman, R. E. (1999). Adult age differences in the functional neuroanatomy of verbal recognition memory. *Human Brain Mapping*, **7**, 115-135.

Nielsen, K. & Peters, A. (2003). The effects of aging on the frequency of nerve fibers in rhesus monkey striate cortex. *Neurobiology of Aging*, **21**, 621-628.

O'Reilly, R.C. & Munkata, Y. 2000. *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press edn.

Peters, A. (2002). Structural changes that occur during normal aging of primate cerebral hemispheres. *Neuroscience and Biobehavioral Reviews*, **26**, 733-741.

Peters, A. & Sethares, C. (2002). The Effects of Age on the Cells in Layer 1 of Primate Cerebral Cortex. *Cerebral Cortex*, **12**, 27-36.

Phillips, W. A. & Silverstein, S. M. (2003). Convergence of biological and psychological perspectives on cognitive coordination in schizophrenia. *Behavioral and Brain Sciences*, **26**, 63-135.

Poe, B. H., Linville, C. & Brunso-Bechtold, J. (2001). Age-Related Decline of Presumptive Inhibitory Synapses in the Sensorimotor Cortex as Revealed by the Physical Disector. *The Journal of Comparative Neurology*, **439**, 65-72.

Reidmiller, M. & Heinrich, B. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. 1, 586-591. 1993. New York, IEEE. IEEE International Conference on Neural Networks (San Francisco).

Ref Type: Conference Proceeding

Schmolesky, M. T., Wang, Y., Pu, M. & Leventhal, A. G. (2000). Degradation of stimulus selectivity of visual cortical cells in senescent rhesus monkeys. *Nature Neuroscience*, **3**, 384-390.

Singer, W., Engel, A. K., Kreiter, A. K. & Munk, M. H. (1995). Visual feature integration and the temporal correlation hypothesis. *Rev Neurosci*, **1**, 252-261.

Spear, P. D. (1993). Neural bases of visual deficits during aging. *Vision Research*, **33**, 2589-2609.

Treisman, A. (1999). Solutions to the binding problem: progress through controversy and convergence. *Neuron*, **24**, 105-110.

Watt, R. J. & Phillips, W. A. (2000). The function of dynamic grouping in vision. *Trends in Cognitive Sciences*, **4**, 447-454.

Wilson, H.R. 1998. In: *Cerebral Cortex: Models of Cortical Circuitry* (Ed. by P.S. Ulinski & E.G. Jones), pp. 14. New York, Plenum.